

# Computergestützte Analyse von NMR- Spektren

## Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)  
der naturwissenschaftlichen Fakultät III  
- Biologie und vorklinische Medizin -  
der Universität Regensburg

durchgeführt am Institut für Biophysik und physikalische Biochemie  
der Universität Regensburg  
unter Anleitung von  
Prof. Dr. Dr. Kalbitzer



vorgelegt von

**Renate Kirchhöfer**

aus Kelheim

April 2005

# Inhaltsverzeichnis

1. EINLEITUNG	1
2. GRUNDLAGEN	3
2.1. NMR- Spektroskopie	3
2.1.1. Physikalische Grundlagen	3
2.1.2. Herkömmliche Methoden	5
2.1.3. Neuere Entwicklungen	6
2.2. Strukturaufklärung	8
2.2.1. Auswertung der Spektren mit AUREMOL	8
2.2.2. Strukturberechnung mit CNS	12
2.3. Beurteilung von Strukturen	14
3. VERBESSERUNG VON STRUKTUREN	17
3.1. Stereospezifische Zuordnung	17
3.2. Das Programm STEREO	19
3.2.1. Funktionsweise	19
3.2.2. Datenbasis	20
3.2.3. Programmbeschreibung und -ablauf	26
3.3. Ergebnisse	33
3.3.1. Programmergebnis	33
3.3.2. Tests und Anwendungsbeispiele	35
3.3.3. Stereospezifische Zuordnung in Datenbanken- eine Statistik	48
3.4. Zusammenfassung und Ausblick	52
3.4.1. Zusammenfassung	52
3.4.2. Schematischer Programmablauf	54
3.4.3. Ausblick	54
4. AUTOMATISCHE ZUORDNUNG	56
4.1. Theoretische Betrachtungen, Probleme	56
4.1.1. Sinn und Zweck	56
4.1.2. Probleme	57
4.2. Das Programm AUNOAS	57
4.2.1. Funktionsweise	57
4.2.2. Datenbasis	59
4.2.3. Programmbeschreibung und -ablauf	60
4.2.4. Programmparameter	71

---

4.3.	Ergebnisse	78
4.3.1.	Ausgabe- Files	78
4.3.2.	Testergebnisse	80
4.3.3.	Stabilität	83
4.3.4.	Anwendungsbeispiele	85
4.4.	Zusammenfassung und Ausblick	85
4.4.1.	Funktionalität	85
4.4.2.	Verbesserungsmöglichkeiten und Ausblick	86
5.	TEMPERATURREIHE VON SARATIN	89
5.1.	Daten	89
5.1.1.	Erzeugung	89
5.1.2.	Aufbereitung	90
5.2.	Automatische Zuordnung mit AUNOAS	91
5.3.	Strukturbestimmung	94
5.4.	Ergebnisse	98
5.4.1.	Räumliche Strukturen	98
5.4.2.	$H^N$ - Shifts	105
5.5.	Ausblick	120
6.	DISKUSSION UND AUSBLICK	121
7.	ZUSAMMENFASSUNG	128
8.	LITERATUR	131
9.	ANHANG	137
9.1.	Vergleich der mittleren Strukturen	137
9.2.	Verlauf der $H^N$ - Shifts über den Temperaturbereich	139
9.3.	Untersuchungen zu unterschiedlichen $T_0$ Werten	151

---

---

Promotionsgesuch eingereicht am 04.04.2005  
Die Arbeit wurde angeleitet von Prof. Dr. Dr. Kalbitzer

Prüfungsausschuss:

Vorsitzender:	Prof. Dr. Witzgall
1. Gutachter:	Prof. Dr. Dr. Kalbitzer
2. Gutachter:	Prof. Dr. Lang
Drittprüfer:	Prof. Dr. Sterner

# 1. Einleitung

Die Aufklärung von Proteinstrukturen hat die biologische Forschung in den letzten Jahrzehnten revolutioniert. Mit dem Wissen um die dreidimensionale Organisation dieser Moleküle wurden viele Reaktionsmechanismen und Funktionsweisen einsichtig, was dazu führte, das noch in den Anfängen stehende Verständnis für die Reaktionsprinzipien in der Zelle zu verbessern.

Nach wie vor steigt jedes Jahr die Zahl der aufgeklärten Strukturen von biologischen Makromolekülen an, jedoch nicht in dem Maße, wie neue Proteine gefunden und entschlüsselt werden. Ein Ende dieser auseinanderklaffenden Schere ist bislang nicht abzusehen, da wegen des damit verbundenen hohen menschlichen Arbeitsaufwands auch weiterhin nur ein sehr geringer Prozentsatz der bekannten Proteine strukturell aufgelöst wird. Außerdem ist mit dem Ende des „Human Genom Projektes“ bei dem das menschliche Erbgut entschlüsselt wurde, ein neuer, gigantischer Pool von Primärstrukturen codierter Proteine entstanden, deren Struktur und Interaktion mit anderen Proteinen geklärt und verstanden werden muss, um weitergehende Erkenntnisse über die Funktion des Lebens an sich zu erhalten.

Daher ist die schnelle und rationelle Aufklärung und Darstellung der räumlichen Strukturmerkmale eines Proteins, insbesondere auch die Veränderung derselben bei Interaktionen mit anderen Proteinen, Ziel vieler wissenschaftlicher Bemühungen.

Eine Proteinstruktur ist kein fester, unverrückbarer Zustand, das Protein muss, ganz im Gegenteil, z.B. bei der Bindung eines Liganden oder bei der Übertragung eines zellulären Signals oftmals ihre Konformation verändern. Die Proteinkonformation mit der geringsten Energie ist die stabilste und auch diejenige mit der größten Anzahl an schwachen Wechselwirkungen. Ein Effekt, den man sich in der NMR bei Energieminimierungsrechnungen und Moleküldynamik-Läufen zunutze macht (vgl. unten).

Auch bei einem Ensemble von gut gefalteten Proteinen in identischer chemischer Umgebung wird nie ein Molekül völlig einem anderen gleichen, außerdem befindet sich jedes System in ständiger Bewegung. Ebenso liegt ein Protein bei unterschiedlicher Umgebung (Temperatur, Druck, pH- Wert) oftmals auch in mehr oder weniger unterschiedlicher Struktur vor, was wiederum die Dynamik biologischer Prozesse demonstriert. Diese kurze Überlegung verdeutlicht schon, welche Anstrengungen bewältigen sind, um Einsichten allein über ein Protein bei bestimmten Zuständen oder während einer Zellreaktionskaskade zu gewinnen.

Nichtsdestoweniger ergeben sich durch umfangreiche Messungen und die Fülle an dadurch erhaltenen Informationen im Zusammenspiel mit diversen Beurteilungskriterien gesicherte Erkenntnisse über das Aussehen von Proteinen in augenfälligen Merkmalen der Strukturelemente, wie dem Zusammenspiel von Helix und Faltblatt und deren Veränderung z.B. bei Ligandenbindung oder veränderter Temperatur.

Zur Zeit gibt es hauptsächlich zwei Methoden, die in der Lage sind, die räumlichen Beziehungen in einer Proteinstruktur mit atomarer Auflösung aufzuklären.

Bei der Röntgenkristallographie werden die Proteine zunächst kristallisiert und anschließend analysiert. Da die Kristallisation aber eine sehr hohe Konzentration an Protein erfordert, das weder aggregiert, noch ausfällt, und immer noch auf rein empirischen Vorgehensweisen beruht, versagt sie bei manchen Proteinen. Des weiteren können bei dem unphysiologischen Prozess der Kristallbildung zusätzlich auch Änderungen in der untersuchten Struktur auftreten. Allerdings ist der Schritt zur räumlichen Struktur, ist die Kristallbildung erst einmal erfolgt, nicht mehr sehr weit. Deshalb sind trotz dieser Einschränkungen über 90% der bekannten Strukturen mit dieser Methode aufgeklärt worden.

Bei der NMR-Spektroskopie werden die Proteine nativ in ihrer natürlichen Umgebung mit Hilfe sehr starker Magnetfelder analysiert. Auch hierbei ist eine hohe Proteinkonzentration notwendig. Bei dieser Art der Analyse fallen große Datenmengen an, die von Menschen ausgewertet werden müssen, außerdem enthalten diese sowohl Artefakte als auch Rauschen und fehlende Signale. Bisher ist es noch nicht zufriedenstellend gelungen diesen Prozess der Spektralanalyse zu automatisieren.

Diese Tatsache ist auch der Grund, warum mit Hilfe der NMR-Spektroskopie deutlich weniger Strukturen aufgeklärt werden als mit der Röntgenkristallographie. Verständlich wird dies, wenn man bedenkt, dass selbst eine geschulte Person oft über ein Jahr benötigt, um eine NMR-Struktur eines Proteins zu erstellen.

In diesem Zusammenhang ist das Hauptziel der vorliegenden Arbeit, sowohl die Grundlagen für die Zuordnung von NMR-Spektren zu verbessern, als auch eine Softwarebasis zu erstellen, die eine weitestgehend automatische Zuordnung von NOESY Spektren erlaubt.

Im ersten Teil soll die Arbeit an einem bereits bestehenden Programm zur stereospezifischen Zuordnung von NMR- Signalen weitergeführt werden. Datenbankanalysen zu diesem Thema vervollständigen und verfeinern die gewonnenen Erkenntnisse.

Im zweiten Teil soll ein Programm entwickelt werden, mit dessen Hilfe NMR-Spektren (NOESY 2D und 3D) schnell und zuverlässig automatisch zugeordnet werden können.

Diese beiden Arbeitsschwerpunkte dienen dazu, den aufwendigen Prozess der NMR-Strukturfindung zu automatisieren und zu beschleunigen.

Die entwickelte Software soll abschließend bei der Messung und Auswertung einer Temperaturreihe des Proteins Saratin eingesetzt und validiert werden. Dabei können strukturelle Untersuchungen an diesem Protein, das über einen weiten Temperaturbereich stabil ist, durchgeführt werden.

Aus dem strukturellen Verhalten dieses Proteins lassen sich Auswirkungen thermischer Einflüsse auf Proteine in einem weiten Temperaturbereich erforschen. Untersuchungen dieser Art fanden bislang nicht statt, da der Arbeitsaufwand für die Bestimmung von Strukturen eines Proteins bei mehr als zwei bis drei verschiedenen Temperaturen mit herkömmlichen, manuellen Methoden kaum realisierbar ist.

Der Wert solcher Untersuchungen zum Temperaturverhalten von Proteinen ist jedoch aus wissenschaftlicher Sicht sehr hoch einzuschätzen, da sie es eventuell ermöglichen könnten, weitere Einblicke in die Dynamik und das Faltungsverhalten von Proteinen zu erlangen.

## 2. Grundlagen

### 2.1. NMR- Spektroskopie

#### 2.1.1. Physikalische Grundlagen

*Die Wirkungsweise und Funktion biologischer Makromoleküle erschließt sich über die Kenntnis ihrer Struktur und Dynamik unter physiologischen Bedingungen.*

Mit diesen Worten könnte man das zentrale Dogma der Molekularbiologie zusammenfassen, dessen Gültigkeit durch zahlreiche Untersuchungen an biologischen Systemen untermauert wurde. Eine ganze Reihe von Stoffwechsel- und Signaltransduktionsprozessen konnte in den letzten Jahren- zumindest teilweise- auf molekularer Ebene erklärt werden. Zu diesem Verständnis haben eine Vielzahl sich gegenseitig ergänzender Methoden beigetragen, von denen die hochauflösende Kernspinresonanz- Spektroskopie (NMR) eine wichtige Rolle einnimmt.

Anhand einiger Schlüsselbegriffe aus obigem Zitat lässt sich die Bedeutung dieses Verfahrens wie folgt einsichtig machen:

- Die NMR- Spektroskopie eignet sich zur Untersuchung biologischer Makromoleküle, weil durch das Phänomen der chemischen Verschiebung (vgl. unten) charakteristische Resonanzfrequenzen für jeden Kern beobachtet werden. Man erreicht dadurch im Prinzip „atomare Auflösung“
- Durch die Messung geeigneter Wechselwirkungen zwischen den magnetischen Kernmomenten erhält man Informationen über interatomare Abstände, Bindungswinkel und –orientierung. Man kann deshalb auf die Struktur des untersuchten Moleküls zurückschließen.
- Im Gegensatz zur Röntgenstrukturanalyse stellt die NMR- Spektroskopie Methoden bereit, die Dynamik von Molekülen innerhalb eines sehr großen Zeit- bzw. Frequenzfensters ( $10^{-4}$  Hz bis  $10^{10}$  Hz) zu studieren
- Es ist mittels Kernspinresonanz möglich, Biomoleküle in gelöster Form und unter nahezu physiologischen Bedingungen (pH- Wert, Ionenkonzentration) zu messen. Man vermeidet damit auch Kristallisationsartefakte, die neben dem partiellen Einfrieren dynamischer Vorgänge bei der Röntgenstrukturanalyse nie ganz ausgeschlossen werden können.

Einschränkungen erfährt die NMR vor allem durch ihre relativ niedrige Empfindlichkeit und den teilweise sehr hohen Auswertungsaufwand. Eine Voraussetzung für erfolgreiche Messungen besteht in der Verfügbarkeit von hochkonzentrierten Proteinproben (millimolarer Bereich), die über die Dauer der Messzeit (Minuten bis Tage) chemisch stabil sein müssen.

Das grundlegende Prinzip aller NMR- Experimente ist die Detektion des Signals der magnetischen Kernmomente  $\vec{\mu}_I$ , die mit dem Kerndrehimpuls (Spin)  $\vec{I}$  verknüpft sind.

$$\vec{\mu}_I = \hbar \gamma_I \vec{I}$$

*Gleichung 1*

( $\gamma_I$ : gyromagnetisches Verhältnis; eine Konstante für jede Kernsorte)

Um eine makroskopische Magnetisierung der Probe zu erreichen, ist das Anlegen eines möglichst starken statischen Magnetfeldes  $\vec{B}_0$  erforderlich. Die Kerne lassen sich dann durch Einstrahlung elektromagnetischer Hochfrequenzpulse (HF- Pulse) resonant anregen. Von zentraler Bedeutung ist die sogenannte Larmorfrequenz  $\omega_L$ , welche die Resonanzfrequenz der Kernspins angibt. Sie ist dem äußeren Magnetfeld proportional.

$$\omega_L = \gamma_I \cdot \|\vec{B}_0\|$$

*Gleichung 2*

Die NMR- Experimente gewinnen ihre Aussagekraft erst durch die Tatsache, dass charakteristische Unterschiede in den Resonanzfrequenzen einzelner Kerne bestehen. Diese „chemische Verschiebung“ (chemical shift) wird durch die partielle Abschirmung des äußeren Feldes durch lokale Elektronendichten verursacht. Am Ort des Kerns wirkt daher nicht die Feldstärke  $\vec{B}_0$ , sondern das effektive Feld

$$\vec{B}_{eff} = (1 - \sigma) \vec{B}_0$$

*Gleichung 3*

Durch die Beobachtung und Auswertung dieser individuell unterschiedlichen chemischen Verschiebungen  $\sigma$  in einem Molekül werden letztendlich Aussagen über die räumliche Struktur möglich.

Aufgabe der NMR- Spektroskopie ist zunächst die Zuordnung der Signale zu den einzelnen Kernen und schließlich die Auswertung von Wechselwirkungen zwischen den einzelnen Spins, in denen die biologisch interessante Information steckt. Fortschritte im Gebiet der Kernresonanz- Spektroskopie ergeben sich daher neben technischen Verbesserungen (Übergang zu stärkeren Magnetfeldern, größere Datenspeicherungs- und Rechnerkapazitäten usw.) vor allem durch die immer besseren Mess- und Auswertverfahren (Pulssequenzen, spezielle Präparation der Proben, Programme zur Automatisierung einzelner Auswertschritte etc.). Die Grenze des Machbaren wird dabei ständig neu definiert. Die folgenden Abschnitte geben einen kurzen Überblick über die wichtigsten herkömmlichen



und aktuellen Techniken. Für detailliertere Informationen wird auf die einschlägige Literatur verwiesen [Ern87, Hau89, Cav95].

### 2.1.2. Herkömmliche Methoden

Der erste Nachweis eines NMR- Signals gelang zwar schon 1945, es war jedoch ein weiter Weg bis zu den heute gängigen Anwendungen, die in zahlreichen wissenschaftlichen Disziplinen routinemäßig genutzt werden.

Als ein wichtiger Meilenstein in der Anwendung kernresonanzspektroskopischer Methoden zur Untersuchung biologischer Proben erwies sich die Entwicklung der gepulsten Fourier-Spektroskopie durch den Nobelpreisträger Richard Ernst im Jahre 1966 [Ern66]. Neben einer Verkürzung der Messzeiten ebnete dieses Verfahren den Weg für die breitbandige Anregung von Spinensembeln und die Entwicklung immer komplexerer Pulsfolgen. Zahlreiche der „klassischen“ Experimente finden auch heute noch ihre Anwendung. Beispiele hierfür sind die Bestimmung von  $T_1$ - und  $T_2$ - Relaxationszeiten und (in Verbindung mit magnetischen Gradientenpulsen) Diffusionsmessungen zur Bestimmung der Molekülgröße.

Zweidimensionale Aufnahmetechniken [Jee71] und die Entwicklung spezieller homo- und heteronuklearer Pulssequenzen schufen schließlich die methodische Grundlage für die hochauflösende Flüssigkeits- NMR- Spektroskopie [Ern87]. Erstmals machte es Sinn, makromolekulare Einheiten mit Hilfe der NMR zu studieren. Dazu mussten Strategien zur Auswertung und Spinsystemzuordnung gefunden werden, um den hohen Informationsgehalt erschließen zu können, der sich in den Spektren großer Proteine verbirgt [Wut86].

Mittlerweile ist die Aufnahme von drei- und vierdimensionalen Spektren zum Standard für die Strukturaufklärung geworden [Gri87, Kay90]. Neben dem Vorteil einer höheren Signaldispersion erreicht man damit vor allem eine weitere Erhöhung des Informationsgehaltes. Durch die geeignete Entwicklung der Magnetisierung können beispielsweise die Extradimensionen dazu genutzt werden, die Signale anhand der Kopplung an benachbarte Spins aufzutrennen („editierte Methoden“). Insbesondere im Zusammenhang mit der Möglichkeit, durch Markierung eines Proteins mit den Isotopen  $^{13}\text{C}$  und  $^{15}\text{N}$ , die wie das Proton  $^1\text{H}$  den Spin  $I = 1/2$  besitzen, auch andere Kerne für die NMR zugänglich zu machen, erlangt dieses Verfahren eine große Bedeutung. Je nach Art des Experiments kann man verschiedene Spin- Spin- Wechselwirkungen selektieren und erhält am Ende einen Satz sich von der Information her ergänzender Messungen.

Es ist klar, dass durch die NMR- Spektroskopie nur in indirekter Weise strukturelle Informationen erhalten werden können. Eine direkte Ortsauflösung auf atomarem Maßstab wäre schon durch die Wellenlänge der eingestrahlten Hochfrequenz (im Bereich eines halben Meters) gar nicht zu erwarten. Stattdessen nutzt man Spin- Spin- Wechselwirkungen wie den Kern- Overhauser- Effekt (NOE- Effekt), um z. B. Abstandsinformationen oder Torsionswinkel zu erhalten. Ziel ist die Bestimmung ausreichend vieler Einschränkungen (restraints), um die dreidimensionale Struktur des Moleküls daraus ableiten zu können. Eine

solche Strukturbestimmung ist sehr rechenintensiv und erforderte die Entwicklung spezieller Methoden zur Simulation der Moleküldynamik [Kap85].

Mit Hilfe des Relaxationsmatrixformalismus kann man NMR- Spektren, ausgehend von berechneten Molekülstrukturen, simulieren und mit den experimentellen Daten vergleichen [Kee84, Goe99b]. Damit erhält man wertvolle Hinweise für die Zuordnung und auch einen Qualitätsmaßstab für die gerechneten Strukturen.

Im Rahmen der pharmakologischen Wirkstoffentwicklung („drug design“) steht oftmals die Suche nach Liganden, die an ein Zielmolekül binden, im Mittelpunkt. Der Einsatz von NMR-Methoden erlaubt sowohl eine schnelle Verifizierung von Komplexbindungen (Linienverbreiterung) als auch die Identifizierung der an der Wechselwirkung beteiligten Reste (Änderung chemischer Shifts, Transfer- NOE- Effekte). Im Prinzip lassen sich sogar Screeninguntersuchungen durchführen, da im Gemisch mehrerer Substanzen die bindende Komponente durch das Verschwinden ihrer scharfen Linien gekennzeichnet ist [Mau00]. Die entsprechenden Untersuchungstechniken sind unter dem Schlagwort „SAR by NMR“ (Structure Activity Relationship by NMR) bekannt geworden.

Wie wir im Rahmen dieser Arbeit zeigen konnten, ist es unter geeigneten Randbedingungen sogar möglich, Proteingemische in einer einzelnen Probe zu trennen, was zur Qualitätssicherung bei der Herstellung von pharmazeutischen Molekülen verwendet werden kann.

Zur Bestimmung der Moleküldynamik existieren mittlerweile theoretische Ansätze, die im Prinzip die Unterscheidung verschiedener Bewegungsvorgänge auf unterschiedlichen Zeitskalen ermöglichen [Lip82, Pal97]. Experimenteller Ausgangspunkt für diese Untersuchungen ist die Messung verschiedener Relaxationsparameter einzelner Spins. Die Kenntnis der dynamischen Vorgänge innerhalb eines Proteins trägt oft zu einem tieferen Verständnis seiner Funktion und grundlegender Wirkungsmechanismen bei.

### 2.1.3. Neuere Entwicklungen

Die aktuellen Bestrebungen im Bereich der Kernresonanz- Spektroskopie laufen einerseits darauf hinaus, die technischen Gegebenheiten zu verbessern, andererseits aber auch neue Arten von Experimenten zu ersinnen, mit deren Hilfe Einschränkungen überwunden und neuartige Effekte zur Informationsgewinnung ausgenutzt werden können. Parallel dazu ist ein wichtiges Feld die Entwicklung von computergestützten Automatisierungsmechanismen, und speziellen Programmen, die dazu beitragen, die äußerst zeitaufwendige Arbeit der Datenauswertung zu minimieren.

Beispielhaft für die erste Kategorie ist die Entwicklung von Kryo- Probenköpfen, welche die Empfängerspulen und den Vorverstärker bis auf 20 Kelvin kühlen, um das thermische Rauschen der Elektrik gering zu halten. Allein durch diese Maßnahme kann das Signal/ Rausch- Verhältnis um einen Faktor von 2- 4 und mehr verbessert werden.

Eine der gravierendsten Einschränkungen für die NMR ist zweifellos die Größe der untersuchten Proteine. Das liegt zum einen an der Signalüberlagerung in den Spektren, zum

anderen nimmt mit zunehmender Molekülmasse die Rotationsdiffusion innerhalb der Lösung ab. Dadurch relaxiert die transversale Magnetisierung schneller, was zur Verbreiterung und Abschwächung der detektierten Kernsignale führt.

Durch vollständige Deuterierung der Proben erreichte man bereits eine deutliche Zunahme der Spin- Relaxationszeiten, was die Untersuchung größerer Moleküle ermöglichte. Einen wesentlichen Durchbruch erbrachte allerdings erst die Entwicklung der TROSY- Technik [Per97]. Im TROSY- Experiment wird selektiv eine Komponente der Multiplettaufspaltung detektiert. Man kann zeigen, dass deren Linienbreite unter bestimmten Umständen fast unabhängig von der Molekülmasse ist. Auf Kosten der Intensität erhält man somit selbst bei großen Proteinkomplexen ausreichend schmale Signale. Die obere Grenze für eine aussichtsreiche Strukturbestimmung hat sich damit von 30 kDa (ca. 270 Aminosäuren) auf bis zu 200 kDa (1800 Aminosäuren) erhöht.

Unter den Verfahren, die grundsätzlich neue Informationen für die Strukturrechnung erbringen, ist vor allem die Messung anisotroper magnetischer Wechselwirkungen zu nennen. Hier ist es an erster Stelle die residuale dipolare Kopplung, welche bereits zur Verfeinerung von Strukturen verwendet wurde, bzw. die es in bestimmten Fällen erst ermöglicht hat, eine hochaufgelöste Struktur zu erhalten [Gro01, Bam01]. Als strukturelevante Parameter erhält man Einschränkungen bezüglich der Ausrichtung chemischer Bindungen in Bezug auf den „Orientierungstensor“ des Moleküls.

Ein wichtiges Feld für zukünftige Entwicklungen stellt die Automatisierung der Auswertung komplizierter NMR- Spektren dar. Die Datenanalyse ist momentan der zeitlimitierende Faktor bei der Entschlüsselung neuer Strukturen. Je nach Größe des untersuchten Proteins und der Güte der aufgenommenen Spektren kann sie sich über mehrere Jahre hinziehen. Daher ist das Interesse an Computerauswertungen und Automatisierungsprozessen stetig gestiegen. Entsprechend gibt es zahlreiche Ansätze, die Arbeitsschritte von der Signalzuordnung bis zu iterativen Strukturrechnungen unter Einsatz von Rückrechnungs- Methoden von spezieller Software durchführen zu lassen [Hel00, Gro02, Oez02, Her02a, Her02b, Hit03, Pri01].

Durch die vollständige Entschlüsselung des menschlichen Erbguts im Rahmen des Human Genome Projects wird in naher Zukunft eine Vielzahl neuer Proteine der Strukturaufklärung zugeführt werden. Mit den bisherigen, also größtenteils manuell gestützten, Auswerteverfahren alleine wird es nicht möglich sein, diese Datenmengen zu bewältigen. Es wird folglich von den weiteren Entwicklungen im Bereich der Automatisierung abhängen, ob die NMR bei den Herausforderungen, die das neue Gebiet der Proteomik mit sich bringt, entscheidenden Beiträge leisten kann.

Ein in diesem Zusammenhang neues und sehr nützliches Programm zur automatischen Zuordnung von NOESY- Spektren wurde im Rahmen dieser Arbeit entwickelt und wird im Abschnitt 1 vorgestellt.

## 2.2. Strukturaufklärung

Der aufwendigste Schritt bei der Bestimmung von Proteinstrukturen mittels NMR-Spektroskopie ist wie bereits erwähnt, die manuelle Auswertung der NMR-Spektren. Die Zuordnungsstrategie für die Resonanzen basiert dabei in erster Linie auf Pulsprogrammen, die skalar gekoppelte, also durch wenige Bindungen getrennte, Spinsysteme korrelieren.

Für die Strukturbestimmung selbst sind momentan noch Abstandsbeschränkungen von dominierender Bedeutung, die man aus dem abstandsabhängigen Kern- Overhauser- Effekt (NOE) erhält. Deren Energiebeiträge zur errechneten Struktur werden als Pseudopotentiale in verschiedenen Strukturrechnungsprogrammen mit verschiedenen Algorithmen zur Simulation der Proteinfaltung minimiert.

Wegen der starken Abstandsabhängigkeit ( $\propto r^{-6}$ ) des NOEs liefert diese Methode nur Informationen über Spins, die näher als rund 0,5 nm zusammenliegen. Deshalb werden zusehends globalere Strukturinformationen wie z.B. die Rest- Dipolkopplungen, zur Strukturbestimmung mit herangezogen. Diese können leichter Strukturinformationen liefern, als man sie durch vollständige Zuordnung aller (gewöhnlich mehrerer tausend) NOESY Kreuzsignale erhalten kann. Dadurch können sie auf jeden Fall zur Verfeinerung (Refinement) der Strukturrechnung als zusätzliche Beschränkungen (Restraints) beitragen. Zusätzlich zu Abstandsinformationen sind außerdem Winkelbeschränkungen hilfreich. Die Konformation des Proteinrückgrats wird dabei wegen der mesomeriebedingten Planarität der Peptidbindung von nur zwei Winkeln  $\Phi$  und  $\Psi$  pro Aminosäurerest vollständig definiert, von denen zumindest ersterer experimentell leicht zugänglich ist.

Zuordnung, Strukturrechnung und Verfeinerung sind in der Praxis zusammenhängende, iterative Prozesse, die mittels geeigneter Programme bewerkstelligt werden.

### 2.2.1. Auswertung der Spektren mit AUREMOL

AUREMOL ist ein in Zusammenarbeit mit Bruker Biospin [Bru04] an unserem Lehrstuhl entwickeltes Programmpaket zur Auswertung multidimensionaler NMR-Spektren, das auf der einen Seite die Möglichkeit der manuellen bzw. interaktiven Analyse von NMR-Spektren bietet, auf der anderen Seite seinen Schwerpunkt aber bei der semi-automatischen bzw. vollautomatischen Auswertung hat.

Im Gegensatz zum traditionellen Ansatz, d.h. der NMR-zentrierten Auswertung (bottom-up Strategie), basiert AUREMOL auf einem molekülorientierten Ansatz, der sogenannten top-down Strategie.

Der Vorteil ist, dass man vor der Analyse bereits möglichst viel Wissen sammelt und bei der anschließenden Auswertung nutzt. Man versucht dabei, ausgehend von der bekannten bzw. vermuteten Molekülstruktur, die NMR-Parameter möglichst genau vorherzusagen und als Leitfaden bei der Auswertung der NMR-Spektren zu benutzen.

Das Konzept, das in AUREMOL benutzt wird, ist schematisch in Abbildung 1 dargestellt.

In einem ersten Schritt wird möglichst viel Information über das Molekül bzw. Protein gesammelt, was als Definition des Moleküls umschrieben ist. Dazu gehört die Primärsequenz des Proteins, also die Aminosäuresequenz, die Zusammensetzung der NMR Probe, d.h. aus welchen Bestandteilen sich die Pufferlösung zusammensetzt und externe physikalische Parameter, wie die Temperatur, der Druck oder der pH-Wert während der Messung.

Zusätzlich wird in AUREMOL eine interne Datenbank verwendet, die möglichst viel a priori Wissen über Proteine und NMR enthält. Die Datenbank beinhaltet z.B. die chemische Struktur der Aminosäuren, statistische Erwartungswerte chemischer Verschiebungen aller NMR- aktiven Kerne der Aminosäuren, Definition verschiedener NMR- Experimente und die Definition von Diederwinkeln, um J-Kopplungen berechnen zu können. Diese Datenbank wurde möglichst allgemein gehalten, sodass sie leicht erweitert werden kann. Der Benutzer ist z.B. in der Lage, neue NMR-Experimente zu definieren oder die Datenbank um die chemische Struktur der vier Nukleotide Cytosin, Guanin, Adenin und Thymin zu erweitern, falls man NMR-Spektren von Nukleinsäuren analysieren will.

Mit diesen Informationen lassen sich nun unterschiedliche Funktionen des Programmpaketes ausführen, z.B. eine erste Zuordnung festlegen, die jedem Kern eine chemische Verschiebung mit Hilfe der Datenbank zugeordnet.

NMR-Spektren können mittels einer Startstruktur, die man z.B. durch Homologie Modelling erhalten kann, mit dem Relaxationsmatrix- Formalismus rückgerechnet werden (RELAX). Aus zugeordneten NOESY- Spektren können mit AUREMOL automatisch Restraint- Files erzeugt werden (REFINE), die für externe Strukturberechnungsprogramme wie DYANA [Gue97] oder CNS [Bru98] als Input dienen. Durch Vergleich von rückgerechneten und experimentellen Spektren kann schließlich die Qualität berechneter Strukturen mittels R- Werten bestimmt werden.

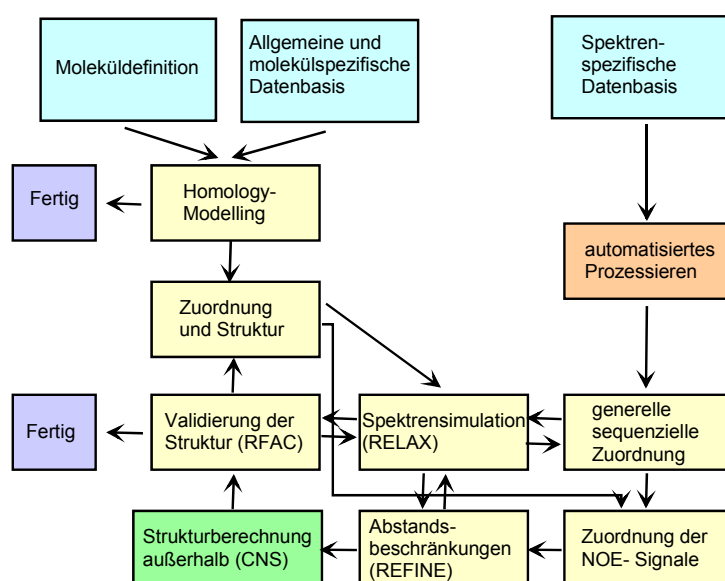


Abbildung 1: Ineinandergreifen verschiedener Programme und Strategien zum Zweck der automatischen Strukturbestimmung [Gro04].

Natürlich können auch experimentelle NMR- Spektren analysiert und einer automatischen Vorverarbeitung unterzogen werden.

Die Hauptfunktionen von AUREMOL beinhalten dabei u.a. die Möglichkeit zum automatischen „Peakpicken“ (mittels lokaler Rausch- Niveau- Berechnung) und Integrieren (Volumenberechnung der Signale) von Spektren und eine Bayes Analyse, mittels derer in einem weiteren Schritt, Rausch- oder Artefaktsignale automatisch entfernt werden können. Für die manuelle Analyse von Spektren steht ein umfangreiches „Strip“- Menü zur Verfügung, mit dem ausgewählte Bereiche der Spektren untersucht werden können. Außerdem können experimentelle Spektren weitgehend automatisch benannt werden. Schließlich stehen auch diverse Programmteile zur Evaluierung und Verbesserung von Strukturen, und eine Möglichkeit für Homologie Modelling zur Verfügung.

Für alle diese unterschiedlichen Funktionen wird eine gemeinsame Datenbasis genutzt, die generelle oder molekülspezifische Informationen enthält, wie oben schon beschrieben wurde.

Im Folgenden werden einige wenige Funktionen kurz näher erläutert, die auch im Rahmen dieser Arbeit Verwendung finden. Für eine detaillierte Übersicht aller Funktionen wird auf das zum Programmpaket gehörende Manual [Gro04], bzw. das Review zum Thema [Gro04b] verwiesen.

### ***Automatisches Peak picken***

Der erste Schritt bei der manuellen und automatischen Analyse ist das sog. automatische Peak picken [Nei95], das alle lokalen Maxima (oder auch Minima, je nach Einstellung des Benutzers) oberhalb eines bestimmten Grenzwertes in einem n-dimensionalen NMR-Spektrum automatisch ermittelt. Diese Maxima, im Folgenden auch kurz Peaks oder Signale genannt, sind Datenpunkte des Spektrums, deren Intensität größer ist als die der nächsten Nachbarpunkte. Der Grenzwert kann dabei manuell vom Benutzer vorgegeben, oder automatisch ermittelt werden.

### ***Wahrscheinlichkeitsberechnung vorhandener Signale***

Die Methode separiert echte NMR-Signale von Rauschsignalen und Artefakten und beruht auf der Annahme, dass verschiedene Klassen (Signal oder Artefakt) verschiedene Verteilungen spezifischer lokaler Merkmale, wie Linienformen, und globaler Merkmale, wie der Symmetrie bezüglich der Diagonalen im Falle von 2D-Spektren, besitzen. Dazu werden, basierend auf dem Bayesschen Theorem [Cor67, Cor69], Wahrscheinlichkeiten dafür berechnet, dass ein Peak zu einer dieser Klassen gehört. Die theoretischen Grundlagen sind in [Ant95] und [Sch97] nachzulesen.

In einem ersten Schritt müssen für die verschiedenen Klassen Wahrscheinlichkeitsverteilungen der verwendeten Merkmale berechnet werden. In der Implementierung von AUREMOL werden zwei Klassen verwendet, eine Klasse echter NMR-Signale und eine Klasse von Rauschsignalen. Die verwendeten Merkmale sind die Signalintensität und 3



verschiedene Volumenverhältnisse [Ant95] als lokale Merkmale, und die Symmetrie bzgl. der Diagonalen im 2D Fall [Sch97] als globales Merkmal.

Dazu müssen Trainingsgebiete für jede Klasse definiert werden, die möglichst nur Peaks dieser Klasse enthalten. Diese Gebiete werden interaktiv vom Benutzer festgelegt.

Mit Hilfe der so erhaltenen Wahrscheinlichkeitsverteilungen kann nun mit dem Entscheidungskriterium von Bayes die Wahrscheinlichkeit  $Q$  berechnet werden, ob ein Peak  $i$  aus einem benutzerdefinierten Testgebiet mit seinen spezifischen Merkmalen zur Klasse der echten NMR-Signalen gehört. Je größer  $p_i$  ist, desto wahrscheinlicher ist Peak  $i$  ein Signal.

Die hier verwendete Methode ist unabhängig von speziellen Linienformen. Charakteristische Eigenschaften von Rauschsignalen und echten Signalen werden mittels iterativer Segmentierung bei unterschiedlichen Segmentierungsgrenzen ausgewählt. Durch eine Analyse dieser Merkmale erhält man Kombinationen, die die bestmögliche Unterscheidung der zwei Klassen von Signalen ermöglichen.

### ***Volumenberechnung mit Fehlerangabe***

Die in AUREMOL integrierte Volumenberechnung analysiert das gegebene Spektrum und berechnet den kleinsten Volumenfehler für jeden Peak auf der Basis von lokalen Rauschlevel- Abschätzungen. Diese Volumenfehler werden dazu verwendet im Programmteil REFINE (siehe unten) Fehlergrenzen für die Abstandsbeschränkungen zu ermitteln. Die Signalvolumina werden automatisch durch iterative Segmentierung des betreffenden Peaks berechnet, wobei die Segmentierung jeweils auf dem höchsten Punkt des Signals beginnt und sich rekursiv nach unten fortsetzt bis Datenpunkte benachbarter Signale oder der Segmentierungsschwellwert erreicht werden.

Dieser Grenzwert, gemessen in Prozent vom Peakmaximum, muss vom Benutzer vorgegeben werden. Zusätzlich muss die maximale Integrationsbreite in Hz für alle Frequenzdimensionen angegeben werden, was sinnvollerweise die maximal zu erwartende experimentelle Linienbreite ist [Gey95].

### ***Rückrechnung mit RELAX***

Ebenso wie die Abstandsinformationen des NOESY Spektrums die dreidimensionale Struktur des Proteins festlegen, bestimmt umgekehrt natürlich auch die Proteinstruktur die Abstandsmatrix der Atome. Infolgedessen lässt sich vorhersagen, welches NOESY Spektrum man für die gegebene Proteinstruktur aus der Zuordnung erwartet. Dies wird mit dem Programm RELAX [Goe94] durchgeführt, wobei z.B. die Mischzeit für den Aufbau des NOEs als Parameter wählbar sind. Chemischer Austausch wird vernachlässigt.

Das Programm gründet dabei auf die Relaxationsmatrixanalyse, die im Gegensatz zum „isolated spin approach“ (ISPA) den Magnetisierungstransfer nicht nur zwischen zwei als isoliert betrachteten Kernen, sondern über Spindiffusionsprozesse zwischen allen Protonen des Systems berücksichtigt.

Die Berechnung der Relaxationsmatrix in Abhängigkeit von den globalen und lokalen Korrelationszeiten erfolgt unter Annahme verschiedener, vom Benutzer wählbarer

Bewegungsmodelle. Im Rahmen des strukturellen Refinements geht man von der Zuordnung der Resonanzen und einer Ausgangsstruktur (oder einem Strukturbündel) aus und rechnet zurück, welche Signale an welcher Position zu erwarten sind. Durch Vergleich mit der Zuordnung der Kreuzsignale im gemessenen NOESY kann man falsch zugeordnete Kreuzsignale identifizieren. Das ist insofern möglich, da experimentelles und simuliertes Spektrum an dieser Stelle nicht zusammenpassen.

Eine weitgehende Übereinstimmung von gemessenem und rückgerechneten Spektrum verifiziert sowohl die Zuordnung als auch Struktur.

### ***Distanzberechnung mit REFINE***

REFINE [Tre04] ist ein in AUREMOL integriertes Tool für die Berechnung der Distanzen zwischen zwei gegebenen Atomen. Es vergleicht benannte experimentelle, und aus der gegebenen räumlichen Struktur simulierte NOE- Signale und gibt Abstandsbeschränkungen für die weitere Verwendung bei der Moleküldynamik- Simulation in für CNS geeignetem Format aus. In einem iterativen Prozess werden hierbei die Raten der Relaxationsmatrix so angepasst, dass die Abweichung zwischen simulierten und experimentellen NOEs minimiert wird. Aus diesen Raten werden dann die entsprechenden Abstandsbeschränkungen ermittelt. Mit den aus diesem Programmteil erhaltenen Informationen ist es möglich, eine Strukturberechnung mit einem externen Programm, z.B. CNS durchzuführen.

## **2.2.2. Strukturberechnung mit CNS**

Der Prozess der Proteinfaltung kann aufgrund des hohen Rechenaufwandes trotz der heute zur Verfügung stehenden Computerkapazitäten nur ansatzweise simuliert werden. Man benötigt deshalb Verfahren, die in „künstlich beschleunigter“ Form zu brauchbaren Ergebnissen führen. Der grundlegende Ansatz dabei ist, die aus den experimentellen Untersuchungen gewonnenen Informationen, die als Einschränkungen (Restraints) des möglichen Konfigurationsraums aufgefasst werden, direkt in die Moleküldynamik- Simulation einzubeziehen. Dies geschieht in der Form von Pseudopotentialen, welche den Faltungsprozess in die gewünschte Richtung lenken.

Bei der Strukturrechnung werden alle empirischen, und durch allgemeine Kenntnisse über Molekülstrukturen gegebenen Restraints (interatomare Abstände, Winkel, Orientierungen, Van-der-Waals-Wechselwirkungen, elektrostatische Potentiale, etc.) kombiniert. Man erhält eine Schar von Strukturen, die anschließend daraufhin überprüft werden, wie gut sie mit den Einschränkungen vereinbar sind. Die Simulation selbst wird im Rahmen eines „simulated annealing“- Protokolls durchgeführt [Kir83], welches im Vergleich zum einfachen Gradientenabstieg mit deutlich höherer Wahrscheinlichkeit Strukturen nahe des globalen Energieminimums findet.

Das verwendete Programm CNS [Bru98] erlaubt die flexible Festlegung von Rechenprotokollen durch Steuerdateien. Zur Simulation benötigt man eine Anfangsstruktur, die aus allen vorhandenen Informationen über die Molekularstruktur, wie Bindungslängen und



Bindungswinkel, generiert wird. Diese Anfangsstruktur kann sowohl ein ausgestreckter Strang ("extended strand") oder eine bereits gefaltete Kette sein.

Ausgehend davon simuliert das Programm die thermische Dynamik der einzelnen Atome und Atomgruppen unter Berücksichtigung der experimentellen Einschränkungen. Ausgangspunkt dafür ist die Integration der Newtonschen Bewegungsgleichungen.

Vor Beginn des Ersten Simulationsschrittes werden allen Atomen der Startstruktur zufällig gewählte Anfangsgeschwindigkeiten aus einer der Temperatur entsprechenden Maxwell-Verteilung zugeordnet. Während der Simulation werden in bestimmten Zeitintervallen (z.B. 0,001- 0,005 ps) die Positionen der Atome berechnet.

In diese Rechnungen gehen sowohl die Startpositionen als auch die mittleren, durch die Maxwell- Verteilung bei einer bestimmten Temperatur gegebenen Geschwindigkeiten ein.

$$F_K(t) = m_K a_K(t) = -grad_K E(r(t))$$

#### Gleichung 4

Wobei  $F_K(t)$  die Kraft auf Teilchen k zum Zeitpunkt t bezeichnet,  $m_K$  die Masse dieses Teilchens und  $a_K$  seine Beschleunigung.

Außerdem enthalten sie auch noch die Beschleunigungen, welche man aus den Newtonschen Bewegungsgleichungen für die angenommenen Kraftfelder ("force fields") erhält, unter deren Einfluss sich die Struktur entwickelt. Das typische molekulare Kraftfeld oder effektive Potential eines solchen Systems lässt sich in empirische und effektive Energieterme trennen [Bru92]:

$$E = E_{emp} + E_{eff} = (E_{Bind} + E_{Winkel} + E_{Diederw} + E_{ungDie} + E_{Ladung}) + (E_{NOE} + E_{Tors})$$

#### Gleichung 5

$E_{emp}$  beinhaltet die ganze Information über die Primärstruktur des Proteins, sowie die Daten über die Topologie und Bindungen des Proteins allgemein. Die Beiträge der kovalenten Bindungen, der Bindungswinkel und der dihedralen Winkel bezüglich  $E_{emp}$  werden näherungsweise durch eine harmonische Funktion beschrieben. Im Gegensatz dazu werden die nicht- kovalenten van- der- Waals Kräfte und die elektrostatischen Wechselwirkungen durch ein Lennard- Jones- oder Coulomb- Potential simuliert.  $E_{eff}$  berücksichtigt die experimentell gefundenen Abstände. Die Winkelzuordnungen und Winkelbeschränkungen werden analog zu den Diederwinkeln durch eine harmonische Funktion angenähert, während  $E_{NOE}$  für solche Entfernungen dann Null ist, wenn sie mit den experimentell gefundenen Werten übereinstimmen. Auch hierfür werden im allgemeinen harmonische Potentiale zugrunde gelegt.

Ziel der Strukturrechnung ist es, eine Konformation des Proteins zu finden, die dem globalen Minimum auf der Energie- Hyperfläche möglichst nahe kommt. Das simulated annealing Verfahren definiert dazu die „effektive Temperatur“  $T$  des Systems. Dieser Parameter entspricht nicht der physikalischen Temperatur, sondern bestimmt die Größe der zu überwindenden Potentialbarrieren. Anfangs wird  $T$  auf einen hohen Wert gesetzt (50000 K), so dass der Konformationsraum recht gleichmäßig abgesucht wird. Im Verlauf der Rechnung senkt man die Temperatur immer weiter ab und erhöht gleichzeitig die Gewichtung der Potentialfaktoren. Auf diese Weise wird der zugängliche Raum immer mehr eingeschränkt, und das System konvergiert im günstigsten Falle stets gegen eine ähnliche Endstruktur.

Das Resultat der Simulation ist eine energieminierte Proteinstruktur. Dabei kann allerdings nicht ausgeschlossen werden, dass statt des globalen nur ein lokales Minimum erreicht wird. Deshalb werden in dieser Arbeit 1000 Strukturen mit zufälligen Anfangsgeschwindigkeiten der Atome berechnet, die dann ihre Endstruktur auf verschiedenen Wegen erreichen. Um die Energien einzelner berechneter Strukturen zu quantifizieren, verwendet die Software CNS empirische Funktionen, welche wie physikalische Potentiale wirken. Damit werden die Beiträge der unterschiedlichen Wechselwirkungen modelliert. Für die NOE-Einschränkungen wird ein „soft- square“ Potential herangezogen, dessen „flacher“ Bereich durch die vorgegebenen Distanzen und deren Fehlergrenzen bestimmt ist. Zu großen Entfernungen hin wächst das Potential nur linear an, damit einzelne unerfüllbare Einschränkungen (z.B. aus fehlerhaften Zuordnungen) nicht dominant werden.

Nach dem Strukturrechenprozess hat man ein Bündel an Strukturen mit minimaler Energie erhalten. Diese sollten sich durch geringe Gesamtenergien im Bereich von ca. 100 kcal/mol auszeichnen, und sich nur sehr wenig voneinander unterscheiden. Die mittlere Abweichung (= root mean square deviation, RMSD) spiegelt hierbei die Präzision der NMR- Strukturen wieder.

Der RMSD- Wert ist unterschiedlich für die verschiedenen Bereiche einer Proteinstruktur. Regionen mit beweglichen Strukturen oder ohne Sekundärstruktur (Loops) zeigen eine größere Abweichung als solche mit einer starren und gut definierten Sekundärstruktur.

Für die im Verlauf der Arbeit durchgeführten Strukturberechnungen wurden das standardmäßig eingestellte Annealing- Protokoll der CNS- Version 1.1, sowie die Defaultwerte für alle Kraftkonstanten verwendet.

## 2.3. Beurteilung von Strukturen

Bei der Betrachtung einer Modellstruktur muss man sich bewusst sein, dass man keinen Blick auf ein wirklich existierendes Molekül wirft, sondern dass diese Struktur nur ein Abbild diverser Experimente, Messungen und Computersimulationen ist.

Wie jedes Modell, kann auch eine NMR- Struktur mehr oder weniger "richtig" sein. Richtig in dem Sinne, dass es die experimentelle Situation gut, sehr gut oder weniger gut beschreibt. Seit es Modelle gibt, hat man versucht, diese auf ihren Wahrheitsgehalt zu überprüfen. Es

gibt sowohl für NMR und "homology modelling" als auch für die Röntgenkristallographie diverse Möglichkeiten die jeweiligen Strukturen zu bewerten.

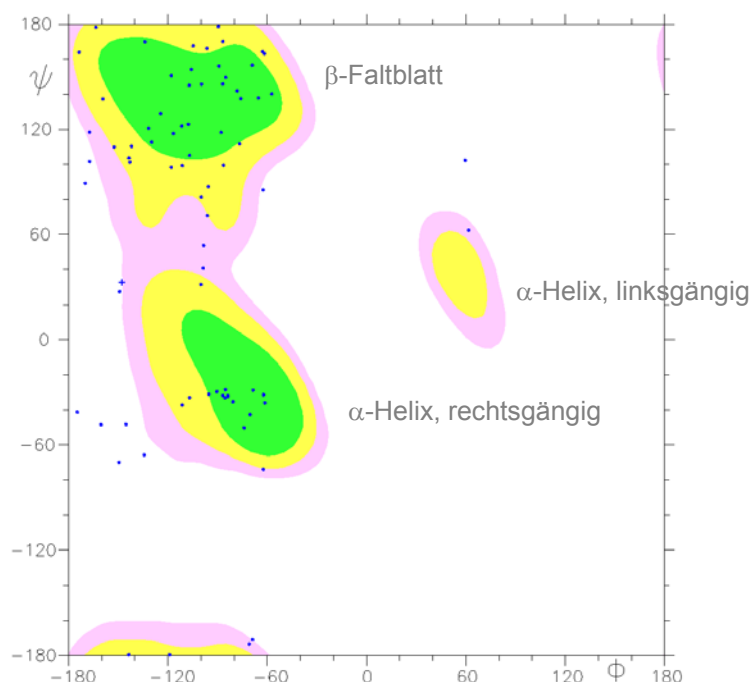
Die Hauptbeurteilungskriterien für NMR- Strukturen lagen bisher überwiegend in Winkel- und Abstandsbeschränkungen, RMSD- Werten und der Güte des Ramachandran- Plots [Las98]. Beschränkungen in der NMR sind z.B. Atomdistanzen und konformationsbestimmende Winkel von NMR Kopplungen (siehe Kapitel 2.1). In der NMR- Strukturbestimmung führt die Konstruktion eines Modells, das mit diesen Beschränkungen in Einklang zu bringen ist, zu einer Struktur, die darüber Aussagen macht, welche Atome durch den Raum benachbart sind.

Um die Güte eines Modells zu bewerten, werden häufig die "restraints per residue" angegeben, also die Gesamtzahl an Beschränkungen für ein NMR- Modell, geteilt durch die Anzahl der Aminosäuren dieses Proteins. Ganz grob gesprochen entspricht ein NMR- Modell mit über 20 restraints per residue in der durchschnittlichen Genauigkeit der Atompositionen einem 0,2 bis 0,25 nm Röntgenkristallographiemodell [Rho99].

Die RMSD- Werte zur mittleren Struktur geben ein Maß dafür, wie sehr die Position jedes Atoms einer Struktur in einem gegebenen Bündel variiert. Der RMSD- Wert für ein Molekül trifft eine Aussage darüber, wie sehr die einzelnen Atome eines gegebenen Bündels von der berechneten mittleren Struktur abweichen.

Die qualitativ besten Modelle ergeben Hauptkettenabweichungen nicht größer als 0,04 nm, mit Seitenkettenwerten um 0,1 nm. Mit diesen Werten kann ein Modell berechnet und gezeichnet werden, das jedes Atom in seiner Durchschnittsposition aufzeigt und die jeweiligen Abweichungen zum Mittelwert aufzeigt (sausage plot). Dieses Modell kann Auskünfte darüber geben, welche Teile der Struktur gut oder weniger gut definiert sind.

In einem Ramachandran- Diagramm werden die Hauptkettenwinkel eines Proteins aufgetragen. Daraus werden Probleme in der Modellstruktur während des Refinement-Prozesses abgelesen. Das aufgetragene Paar aus dem Winkel  $\Phi$  und  $\Psi$  einer einzelnen Aminosäure ist streng durch sterische Behinderungen beschränkt. Die erlaubten Kombinationen sind gefärbt eingezeichnet (siehe Abbildung 2) Jeder Punkt auf dem Diagramm, das erstellt wird, repräsentiert die beiden Diederwinkel  $\Phi$  und  $\Psi$  auf jeder Seite des  $\text{C}\alpha$ - Atoms einer Aminosäure. Ramachandran- Plots sind eine große Hilfe, unrealistische Regionen im Strukturmodell zu finden.



*Abbildung 2: Beispiel für einen Ramachandran- Plot. Es zeigt die theoretisch erlaubten Winkelkombinationen für bestimmte Sekundärstrukturelemente*

In der wichtigsten Konkurrenzmethode, der Röntgenkristallographie verfügt man mit dem kristallographischen R- Faktor zusätzlich über ein echtes Qualitätsmaß für die berechnete Struktur. Dabei vergleicht man experimentelle und rückgerechnete Spektren. Der R- Faktor ist hierbei die mittlere Abweichung der aus der Struktur zurückgerechneten Reflexintensitäten von den gemessenen [Bru87].

In früheren Arbeiten wurde von uns auch für die NMR ein vergleichbarer Qualitätsfaktor eingeführt (RFAC, [Gro00]). Dabei wird ein benanntes experimentelles NMR- Spektrum mit dem zurückgerechneten Spektrum aus einer bestimmten Strukturvorstellung verglichen. Damit können sowohl globale R- Werte (für die ganze Struktur) als auch für bestimmte Teile (z.B. Sekundärstrukturelemente) oder für unterschiedliche Distanzklassen (long- range, short- range NOEs) berechnet werden. Diese R- Faktoren geben dann Aufschluss darüber, wie gut die jeweilige Strukturvorstellung mit den tatsächlichen experimentellen Daten übereinstimmt.

In diesem Zusammenhang haben wir schon erste Versuche zur Entwicklung eines Programmes unternommen, das auf der Basis dieser R- Werte die stereospezifische Zuordnung von Proteinstrukturen überprüft und ggf. berichtigt. Diese Arbeit wurde weitergeführt und um eine Analyse der in einschlägigen Datenbanken vorhandenen stereospezifischen Zuordnungen bereichert (s. Abschnitt 1).

### 3. Verbesserung von Strukturen

#### 3.1. Stereospezifische Zuordnung

Für die Bestimmung von Proteinstrukturen ist es notwendig 2D oder 3D NOESY Spektren auszuwerten und die dort auftretenden Peaks (NOEs) so vollständig, und vor allem so richtig wie möglich zu benennen.

Ein besonderes Problem bei der Zuordnung stellen stereospezifisch zuzuordnende Signale von z.B. Methylenprotonen der einzelnen Aminosäurereste dar.

Bei diesen Paaren von Protonen gestaltet sich die Zuordnung der Resonanzen zu ihren entsprechenden Signalen (im Folgenden als stereospezifische Zuordnung bezeichnet) in der Praxis oftmals recht zeitaufwendig und schwierig, da nicht mit einfachen Mitteln geklärt werden kann, welche Resonanz zu welchem Proton gehört. Eine Möglichkeit ist, J-Kopplungen und NOEs auszuwerten.

Zusätzliche Schwierigkeiten treten durch die häufigen Überlappungen der Resonanzsignale bei größeren Biomolekülen auf. Da jedoch Strukturrechnungen auf der Basis von Abständen, besonders weitreichenden Abständen, basieren, ist es sehr hilfreich und nützlich, diese Zuordnungen richtig zu treffen.

Abbildung 3 zeigt die Situation bei der Seitenkettenzuordnung eines NMR- Experiments. Wenn man die Signale der  $H^{\beta 3}$  und  $H^{\beta 2}$  richtig zugeordnet hat, und damit die Zuordnung der gegebenen reellen Situation entspricht, müsste das Signal vom  $H^{\beta 2}$  zum  $H^N$  der sequenziell entfernten Aminosäure stärker sein.

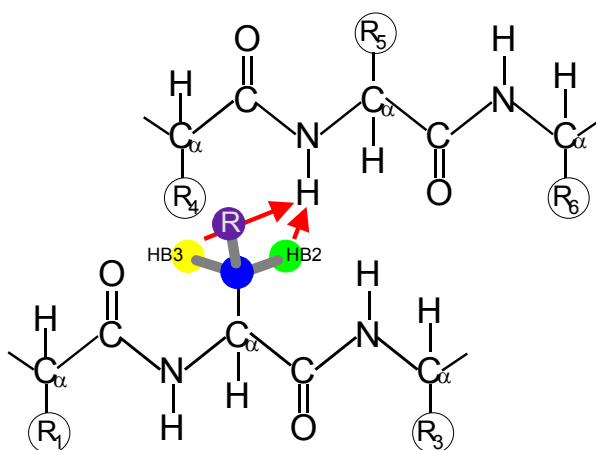


Abbildung 3: Lagebeziehungen von  $H^{\beta}$ - Atomen.  $H^{\beta 3}$  ist weiter vom betrachteten  $H^N$  entfernt als  $H^{\beta 2}$ . Dementsprechend muss das beobachtete NMR- Signal von  $H^{\beta 3}$  weniger intensiv sein.

Hat man aber "falsch" zugeordnet und das stärkere Signal mit  $H^{\beta 3}$  benannt, führt das zu Unstimmigkeiten in den Strukturbeziehungen, da die betrachteten Atome auch noch zu anderen Atomen in Beziehung treten, und Strukturberechnungen auf dieser Basis durchgeführt werden. Im Extremfall führen viele solcher Unstimmigkeiten zu einer zumindest teilweise falsch gefalteten Proteinstruktur.

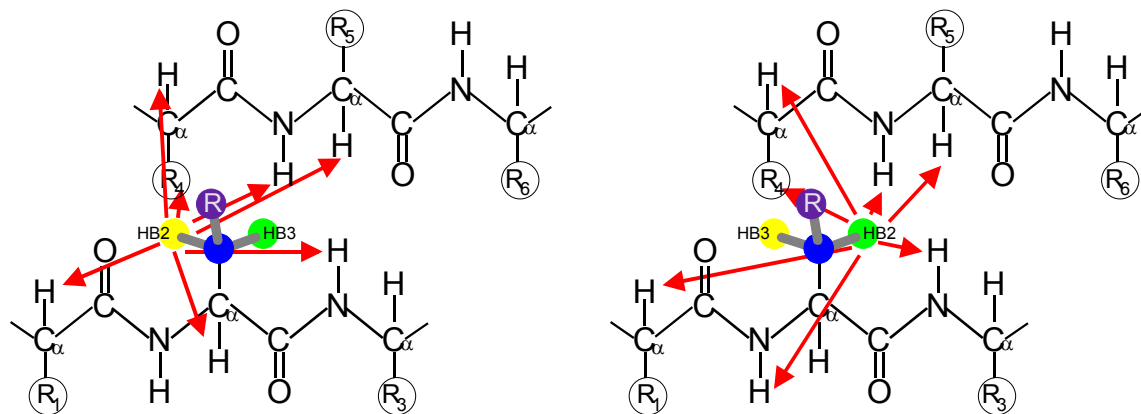


Abbildung 4: Kontakte zu anderen Protonen von  $H^{\beta 2}$  aus, unter der Annahme,  $H^{\beta 2}$  wäre das grüne Atom (links) und wenn man annimmt, das gelbe Atom wäre  $H^{\beta 2}$  (rechts).

Wie Abbildung 4 zeigt, sind die Distanzen zu anderen Protonen von  $H^{\beta 2}$  aus gesehen sehr unterschiedlich, je nachdem, welches H-Atom man als  $H^{\beta 2}$  bezeichnet.

Im NMR-Spektrum selbst ist es jedoch nicht ohne weiteres zu erkennen und zu entscheiden, welches Signal zu welchem  $H^{\beta}$  ( $\gamma$ ,  $\delta$ ,  $\epsilon$  etc.) gehört. Es ist daher noch immer übliche Praxis diese Atome entweder statistisch in  $H^{\beta 3}$  und  $H^{\beta 2}$  einzuteilen, oder gar nicht zu unterscheiden, indem man beide Atome als  $H^{\beta}$  klassifiziert. Dadurch müssen dann in der Strukturrechnung Pseudoatome und/oder größere Abstandsgrenzen eingeführt werden, was wiederum zu weniger gut definierten Strukturen führt.

Zur Zeit existieren verschiedene Programme zur automatischen stereospezifischen Zuordnung. Die meisten davon arbeiten hauptsächlich mit einer kombinierten Analyse aus lokalen Distanzen und Einschränkungen auf der Basis skalarer Kopplungen.

So versucht z.B. STEREOSEARCH [Nil90] eine stereospezifische Zuordnung auf der Basis intraresidueller und sequenzieller NOE-Kontakte und Kopplungskonstanten, die mit den Diederwinkeln mittels der Karplus-Gleichung in Beziehung gesetzt werden.

Dafür wird eine Datenbasis, die Torsionswinkel, zugehörige  $^1\text{H}$ - $^1\text{H}$  Kopplungskonstanten und Abstände zwischen den Protonen enthält, automatisch nach Datensätzen durchsucht, die innerhalb spezifizierter Fehlergrenzen zu den experimentellen Datensätzen passen.

HABAS [Gue89,91], ANGLESEARCH [Pol95] und HYPER [Tej99] benutzen Gitternetz-Suchmethoden. HYPER z. B. berechnet einen Satz aus  $\Phi$ ,  $\Psi$  und  $X_1$  Diederwinkeln und  $\text{C}^{\beta}\text{H}_2$  stereospezifische Zuordnungen, GLOMSA [Gue91] untersucht berechnete Strukturen

auf gleichbleibende Positionen von Prochiralen Gruppen, um Zuordnungen für eine weitere Verbesserung dieser Strukturen zu erhalten.

## 3.2. Das Programm STEREO

### 3.2.1. Funktionsweise

Im Gegensatz zu den Eingangs erwähnten Programmen ist das von uns entwickelte STEREO ein einfach zu bedienendes Programm, das auf dem Vergleich von experimentellen und simulierten  $^1\text{H}$  NOESY NMR Spektren basiert. Bei diesem Vergleich werden zusammengehörende Methylenprotonen (auch Methylgruppen von Valin oder Leucin, oder  $\text{NH}_2$ - Gruppen von Seitenketten der gleichen Aminosäure, wenn sie nicht unterscheidbar sind) identifiziert und ihre korrekte Benennung mittels Berechnung eines NMR R- Faktors überprüft.

Der Grundgedanke zur Überprüfung der stereospezifischen Zuordnung basiert auf der Tatsache, dass bei der Rückrechnung eines Spektrums möglichst alle experimentellen Signale erklärt sein sollten, und auch die Intensitäten der zusammengehörenden Signale- die ja abstandsabhängig sind- sich entsprechen. Tritt nun der Fall eines falsch zugeordneten  $\text{H}^\beta$  auf, so wird- um beim Beispiel aus 3.1 zu bleiben- die Intensität eines zurückgerechneten Signals, an dem  $\text{H}^{\beta 3}$  beteiligt ist, größer sein, als die des entsprechenden Signals von  $\text{H}^{\beta 2}$ , wogegen in der experimentellen Liste die Intensität dieses  $\text{H}^{\beta 2}$  größer ist, als die des  $\text{H}^{\beta 3}$ . Daher wird sich der lokale R-Wert dieser Aminosäure verbessern, wenn man die Benennungen vertauscht.

Dazu haben wir eine Routine entwickelt, die jeweils den R- Wert der entsprechenden Signale im Originalfall und im getauschten Fall berechnet und vergleicht, und die Benennung von z.B.  $\text{H}^{\beta 3}$  und  $\text{H}^{\beta 2}$  dann vertauscht, wenn sich der R- Wert verbessert.

Dieser R- Faktor ist dem R- Faktor aus der Röntgenkristallographie ähnlich und ist dem Programm RFAC in AUREMOL entlehnt (vgl. Abschnitt 2.3). Die in STEREO verwendete Formel zur Berechnung des R- Wertes erfolgt nach

$$R_3(\alpha) = \sqrt{\frac{\sum_{i \in A} (V_{\text{exp},i}^\alpha - sf_\alpha \cdot V_{\text{calc},i}^\alpha)^2 \cdot p_{\text{exp},i}^2}{\sum_{i \in A} V_{\text{exp},i}^{2\alpha} \cdot p_{\text{exp},i}^2}} \quad sf_\alpha = \frac{\sum_{i \in A} (V_{\text{exp},i} \cdot V_{\text{calc},i})^\alpha}{\sum_{i \in A} V_{\text{calc},i}^{2\alpha}}$$

Gleichung 6: R- Wert Berechnung und Berechnung des Skalierungsfaktors im Programm STEREO.

Wobei:  $V_{\text{exp}}$  das Volumen eines experimentellen Signals, und  $V_{\text{calc}}$  das Volumen eines simulierten Signals ist,  $sf$  stellt einen Skalierungsfaktor zwischen simulierten und experimentellen Spektrum dar,  $p$  ist eine gegebene Wahrscheinlichkeit dafür, ob ein Peak

vom zu untersuchenden Protein herrührt (siehe 2.2.1 Auswertung der Spektren mit AUREMOL).

Im Programmablauf haben wir einen Teil des Programms RFAC so abgeändert, dass es intern mit STEREO zusammenarbeitet, und in STEREO integriert.

### 3.2.2. Datenbasis

Wie im einleitenden Kapitel über AUREMOL bereits angedeutet wurde, besteht die interne Datenstruktur des Programms aus zahlreichen Komponenten, die ineinander greifen und optimal auf das Konzept von AUREMOL abgestimmt sind. Im Folgenden sollen die für die Arbeit relevanten Datenformate kurz beschrieben werden.

Diese Datenstruktur enthält die komplette Information, die für die Auswertung der NMR-Spektren von Bedeutung ist, und ist an die Notation des STAR Formats angelehnt [Bmr04]. Das STAR Format wurde von der BioMagResBank als Datenformat für die NMR-Spektroskopie vorgeschlagen. Es findet eine Unterteilung in einzelne Sektionen statt. Jede Sektion beginnt mit einer Definition der verwendeten physikalischen Größen, woran sich zeilenweise die Werte für diese Parameter anschließen. Alle Informationen werden intern im IUPAC- Standard verarbeitet und gespeichert. Zur weiteren Beschreibung der Datenstruktur siehe [Gan02].

#### 3.2.2.1. Verwendete Daten

##### ***Masterliste des experimentellen Spektrums:***

In AUREMOL wurde ein neues Listenformat entwickelt, das die für die Auswertung relevanten Daten aus den Spektren zusammenfasst. Dies umfasst sowohl experimentelle als auch simulierte Spektren. Die Masterliste eines Spektrums enthält wichtige Informationen aller gepickten Peaks. In Abbildung 6 ist ein Beispiel zu sehen.

Es beginnt mit einem Header, der das NMR-Experiment, die Dimension des Spektrums und andere wichtige Prozessierungsparameter, wie z.B. die Spektrenbreite in Hz oder ppm enthält. Dann werden alle gepickten Peaks aufgelistet, beginnend mit dem Peaknamen und evtl. Kommentaren. Die chemischen Verschiebungen in jeder Dimension, die Intensität, das Volumen des Peaks und der Qualitätswert, der angibt, mit welcher Wahrscheinlichkeit es sich um ein echtes NMR-Signal oder um ein Rauschsignal handelt, folgen.

In einer weiteren Zeile werden vorhandene Subpeaks und die weiter oben schon erwähnten berechneten Volumenfehler angegeben.

Die Masterliste enthält somit alle verfügbaren Informationen zum aktuell betrachteten experimentellen/ simulierten Spektrum. Idealerweise sollten alle dort vorkommenden Signale zugeordnet, d.h. benannt sein.

##### ***Peakliste des simulierten Spektrums:***



Das simulierte Spektrum wird unter Einbeziehung der Spindiffusion (siehe 2.2.1 Rückrechnung mit RELAX) aus der vorhandenen Proteinstruktur zurückgerechnet. Ist die erstellte Proteinstruktur perfekt richtig und die Simulation ideal, sind experimentelles und simuliertes Spektrum identisch. Für ein simuliertes Spektrum wird zusätzlich zur Masterliste ein File ausgegeben, das die Namen und Koordinaten der Signale, ihr Volumen und daraus berechnet, die interatomaren Abstände in tabellarischer Form ausgibt, was die Weiterverarbeitung in externen Programmen, z.B. EXCEL möglich macht. Ein Beispiel zeigt Abbildung 7.

### ***Pdb- File:***

Im Pdb- File sind alle Informationen zur aktuell betrachteten Proteinstruktur zusammengefasst. Insbesondere finden sich dort alle vorhandenen Atome des betrachteten Moleküls und deren räumlichen Koordinaten (siehe Abbildung 8). Ein Pdb- File wird bei der Strukturrechnung mit z.B. CNS erhalten und kann nach einer Konvertierung ins IUPAC-Format in AUREMOL verwendet werden.

### ***Compound- File:***

Die Compounddatei wird aus den Parametern, die in der globalen Datenbank enthalten sind, und einer Sequenzdatei erzeugt. Die Sequenzdatei enthält die Primärsequenz des Biomoleküls im Ein- bzw. Drei-Buchstaben-Code.

Die Compounddatei definiert einen Bestandteil der NMR-Probe, was entweder das interessierende Protein sein kann oder ein anderes Molekül der Lösung. In Abbildung 10 ist ein Auszug aus einer solchen Datei zu sehen. Sie gliedert sich in drei Abschnitte bzw. Sektionen.

Im ersten Abschnitt werden alle Atome in sequentieller Reihenfolge spezifiziert. Es beginnt mit der Nummer der Aminosäure in der Sequenz und der Atomnummer in der Aminosäure. Dann folgt der Name der Aminosäure im Drei-Buchstaben-Code, der Atomname im IUPAC-Format und der Atomtyp. Der Atomtyp unterscheidet zwischen  $^1\text{H}$ ,  $^{13}\text{C}$  und  $^{15}\text{N}$  Kernen.

Zusätzlich finden sich Informationen, ob ein Aliasname existiert, zu welchen anderen Atomen dieser Kern magnetisch äquivalent ist, und über die Anisotropiekonstanten der chemischen Verschiebung.

Anschließend wird die chemische Struktur angegeben, d.h. für jedes Atom ist festgelegt mit welchen anderen Atomen es kovalent gebunden ist.

Der letzte Abschnitt definiert Diederwinkel, und zwar sowohl vom Rückgrat als auch von den Seitenketten des Proteins. Zusätzlich sind die Karpluskonstanten A, B und C angegeben.

Anschließend werden für alle Atompaare, die über maximal drei Bindungen miteinander gekoppelt sind, J-Kopplungskonstanten bzw. einer der oben definierten Diederwinkel angegeben, um mit der Karplusgleichung bei der Simulation von NOESY-Spektrum die J-Kopplungsstärke berechnen zu können.

In der Compounddatei sind keinerlei probenspezifische Informationen enthalten, wie z.B. chemische Verschiebungen.

**Meta- File:**

Das Meta- wie auch das Compound- File wird dazu verwendet, Informationen über die Zuordnung zu erhalten.

Die Metadatei fasst alle bisherigen Informationen zusammen. Sie wird aus den einzelnen Compounddateien und der Masterliste erzeugt. Diese besteht wieder aus mehreren Sektionen. In Abbildung 9 ist ein Ausschnitt zu sehen.

Im ersten Abschnitt werden die einzelnen Bestandteile der NMR-Probe angegeben. Es sind die Compounddateien aufgeführt sowie Informationen über die Konzentration der Bestandteile in der Probe und evtl. Isotopenmarkierungen. Anschließend sind einige weitere physikalische Parameter zu setzen, wie z.B. der pH-Wert oder die Temperatur bei der Messung, die bisher bei der Auswertung keine Rolle spielen und daher nur als Defaultwerte gesetzt sind.

Die weiteren Sektionen definieren die einzelnen Bestandteile näher. Zunächst sind Klassen für verschiedene Atomtypen definiert, die z.B. die Bewegungsmodelle für die Simulation der NOESY-Spektren näher beschreiben. Dann werden alle Zuordnungen aus der Masterliste aufgeführt, allerdings nicht mit den entsprechenden Atomnamen, sondern mit zwei Indizes, die auf das korrespondierende Atom in der Compounddatei verweisen.

Darauf folgt der über alle experimentellen Spektren gemittelte Wert für die chemische Verschiebung, und die Genauigkeit mit der dieser Wert gegeben ist.

Weiter ist eine der oben definierten Atomklassen und eine evtl. bestimmte experimentelle Linienbreite angegeben. Daran schließt sich ein weiterer Abschnitt über experimentell bestimmte J-Kopplungen an, die aus der Masterdatei extrahiert wurden. Mit den Angaben, die in der Metadatei und den zugehörigen Compounddateien enthalten ist, ist es nun möglich, NOESY-Spektren zu simulieren, oder auf alle verfügbaren Informationen zu einem benannten Spektrum zuzugreifen.

Speziell im Meta- File werden auch die Ambiguity- Codes zu den zugeordneten Atomen hinterlegt, die Aufschluss über die Eindeutigkeit der Benennung geben.

So ist beispielsweise ein Ambiguity- Code von 1 ein Zeichen dafür, dass die zuordnende Person das Signal eindeutig zugeordnet hat. Alle Codes (1-5 und 9) folgen der Konvention, die in der BioMagRes- Datenbank nachzulesen ist [Bio04].

```
#####
#   Chemical Shift Ambiguity Code Definitions   #
#                                               #
#   Codes           Definition                 #
#                                               #
#   1               Unique                     #
#   2               Ambiguity of geminal atoms or geminal methyl #
#                   proton groups              #
#   3               Aromatic atoms on opposite sides of the ring #
#                   (e.g. Tyr HE1 and HE2 protons) #
#   4               Intraresidue ambiguities (e.g. Lys HG and    #
#                   HD protons)                 #
#   5               Interresidue ambiguities (Lys 12 vs. Lys 27) #
#   9               Ambiguous, specific ambiguity not defined    #
#                                               #
#####
```

Abbildung 5: Ambiguity- Code- Definitionen der BioMagResBank.

Alle Daten können sowohl im CNS Format als auch im IUPAC Standard eingelesen werden. Wichtig ist jedoch, dass sowohl im pdb-, als auch im compound-/meta- und im RELAX-output- File dieselbe Benennungskonvention verwendet wird.

Die interne Datenbearbeitung in AUREMOL erfolgt allerdings, wie bereits erwähnt im IUPAC-Format.

### 3.2.2.2. Format der Eingabedaten

#### **Masterliste des experimentellen Spektrums:**

```
Masterlist (Volumes & intensities rescaled with NC_Proc, scalingfactor sf=2^NC_Proc)
=====
HEADER
  EXPERIMENT:      NOESY->H/H
  DIMENSION:       2
  MIXINGTIME_SEC:  0.15
  ACQUISITIONDELAY: *
  RESONANCEFREQUENCY_1: 800.130005
  RESONANCEFREQUENCY_2: 800.130005
  SPECTRAL_WIDTH_HZ_1: 10197.654007
  SPECTRAL_WIDTH_HZ_2: 10197.654007
  SPECTRAL_WIDTH_PPM_1: 12.744996
  SPECTRAL_WIDTH_PPM_2: 12.744996
  OFFSET_PPM_1:    11.897678
  OFFSET_PPM_2:    11.897678
  FILTER_1:        gaussian
  FILTER_2:        gaussian
  LINEBROAD_HZ_1:  2.500
  LINEBROAD_HZ_2:  2.500
  AQ_MODE_1:        simultaneous
  AQ_MODE_2:        simultaneous
  SEQUENCE:         *
END_HEADER

Total amount of peaks: 13395
Including subpeaks: 231805

PEAKLABEL: HD1 81/HD1 81
COMMENT: Stereo
FEATURES: AMBIGUITY 1/1
COMPOUNDS: 1 1
PEAKDESCRIPTION:
  coord: 1   coord: 2   intensity   volume   quality
  978 -0.27479   1955 -0.26857   1466486   10018257   -1.000
SUBPEAKS: 0
-275 -269 1000

PEAKLABEL: HD1 81/HD3 81
COMMENT: Stereo
FEATURES: AMBIGUITY 1/1
COMPOUNDS: 1 1
PEAKDESCRIPTION:
  coord: 1   coord: 2   intensity   volume   quality
  978 -0.27479   1841 0.44087   189563   1277848   -1.000
SUBPEAKS: 0
-275 441 1000

PEAKLABEL: HD1 81/HD1 8
COMMENT: Stereo
FEATURES: AMBIGUITY 1/1
COMPOUNDS: 1 1
```

*Abbildung 6: Format einer experimentellen Masterliste, generiert von AUREMOL, mit Angabe des Experiments und der Dimension im Header. (Im 2D- Fall üblicherweise bezeichnet als Arr.ml, und lokalisiert im Ordner des verwendeten Spektrums)*

**Peakliste des simulierten Spektrums:**

name 1	name 2	shift 1	shift 2	dist [Å]	rate [Hz]	noe	width 1 [Hz]	width 2 [Hz]
HD1 81	HD1 81	-0.268	-0.268	0.000	6.672	1.180e+000	15.529	15.529
HD1 81	HD2 81	-0.268	0.445	3.337	-0.694	1.505e-001	15.529	15.055
HD1 81	HD1 8	-0.268	0.468	4.717	-0.087	2.704e-002	15.529	15.604
HD1 81	HD1 74	-0.268	0.582	6.378	-0.014	7.298e-003	15.529	16.497
HD1 81	HD1 61	-0.268	0.618	3.720	-0.362	7.011e-002	15.529	15.415
HD1 81	HD1 22	-0.268	0.653	7.179	-0.007	2.511e-003	15.529	15.193
HD1 81	HD2 22	-0.268	0.662	7.052	-0.008	2.752e-003	15.529	15.618
HD1 81	HG2 77	-0.268	0.662	3.258	-0.802	1.550e-001	15.529	15.760
HD1 81	HD1 63	-0.268	0.683	4.117	-0.197	4.437e-002	15.529	16.560
HD1 81	HD1 35	-0.268	0.759	6.235	-0.016	6.464e-003	15.529	14.858
HD1 81	HD1 86	-0.268	0.764	6.295	-0.015	1.185e-002	15.529	15.793
HD1 81	HG2 63	-0.268	0.765	6.931	-0.009	7.284e-003	15.529	15.698
HD1 81	HD2 86	-0.268	0.824	6.811	-0.010	9.310e-003	15.529	16.029
HD1 81	HD2 44	-0.268	0.847	7.054	-0.008	2.452e-003	15.529	15.512
HD1 81	HD1 44	-0.268	0.849	6.792	-0.010	4.502e-003	15.529	15.923
HD1 81	HG1 80	-0.268	0.966	6.239	-0.016	1.111e-002	15.529	14.455
HD1 81	HG13 61	-0.268	0.985	4.849	-0.043	1.070e-002	15.529	11.408
HD1 81	HG13 77	-0.268	1.020	6.098	-0.011	5.201e-003	15.529	8.265

Abbildung 7: Peakliste eines simulierten Spektrums, generiert mit der „backcalculation“- Funktion im „calculation“- Menü von AUREMOL. Nur diese, textbasierte Form wird für STEREO benötigt, es ist also nicht notwendig ein simuliertes Spektrum mit Masterliste zu generieren (zeitaufwendiger!).

**Pdb- File**

ATOM	1	N	MET	1	8.446	12.266	-4.329	1.00	0.00	HPR
ATOM	2	CA	MET	1	7.699	12.341	-5.559	1.00	0.00	HPR
ATOM	3	HA	MET	1	6.709	12.857	-5.455	1.00	0.00	HPR
ATOM	4	CB	MET	1	8.515	13.292	-6.544	1.00	0.00	HPR
ATOM	5	HB3	MET	1	8.611	14.367	-6.175	1.00	0.00	HPR
ATOM	6	HB2	MET	1	8.012	13.382	-7.504	1.00	0.00	HPR
ATOM	7	QB	MET	1	8.311	13.875	-6.839	1.00	0.00	HPR
ATOM	8	CG	MET	1	9.952	12.746	-6.870	1.00	0.00	HPR
ATOM	9	HG3	MET	1	9.812	11.962	-7.586	1.00	0.00	HPR
ATOM	10	HG2	MET	1	10.299	12.244	-5.983	1.00	0.00	HPR
ATOM	11	QG	MET	1	10.055	12.103	-6.784	1.00	0.00	HPR
ATOM	12	SD	MET	1	11.043	13.947	-7.593	1.00	0.00	HPR
ATOM	13	QE	MET	1	10.319	13.464	-9.648	1.00	0.00	HPR
ATOM	14	CE	MET	1	10.453	13.562	-9.280	1.00	0.00	HPR
ATOM	15	HE1	MET	1	11.203	13.969	-10.073	1.00	0.00	HPR
ATOM	16	HE2	MET	1	9.477	13.980	-9.460	1.00	0.00	HPR

Abbildung 8: Pdb- File einer gegebenen Proteinstruktur. In dieser Datei sind alle Atome eines Moleküls mit ihren dreidimensionalen Koordinaten aufgeführt. Da im Pdb- File alle Atome vollständig aufgeführt sind<sup>1</sup>, stellt es die Grundlage für jede weitere Informationsbeschaffung und Datenaufbereitung dar.

<sup>1</sup> Im Meta-/ Compound- File sind nur die jeweils sequenziell zugeordneten Atome aufgeführt.

## Meta- File

```

SHIFTS:
_Residue_seq_code
_Atom_num_code
_Atom_alias
_Chem_shift_value
_Chem_shift_value_error
_Chem_shift_ambiguity_code
_Atom_class
_Linewidth
1  4 - 4.68000    0.20  1  0  -
1  9 - 2.06400    0.20  1  7  -
1  8 - 2.19600    0.20  1  7  -
1 15  HE 2.04731    0.20  1  1  -
1 12 - 2.52276    0.20  1  7  -
1 11 - 2.61217    0.20  1  7  -
2  4 - 4.72900    0.20  1  0  -
2  9 - 1.59000    0.20  1  7  -
2  8 - 2.08582    0.20  1  7  -
2 12 - 2.29500    0.20  1  7  -
2 11 - 2.37300    0.20  1  7  -

```

Abbildung 9: Meta- File eines experimentellen Spektrums. In ihm sind u.a. Atomnummern basierend auf Aminosäuren, chemische Verschiebungen und Ambiguity- Codes der zugeordneten Signale verzeichnet.

## Compound- File

```

COMPOUND: hpr
section_sequenzdefinition
_Residue_seq_code
_Atom_num_code
_Residue_label
_Atom_name
_Atom_type
_Atom_alias
_Atom_equivalent
_Atom_CSA
1  1 MET  HN H  -  -  8.95
1  2 MET  N N  -  - 157.00
1  3 MET  CA C  -  -  40.00
1  4 MET  HA H  -  -  8.95
1  5 MET  C C  -  -  40.00
1  6 MET  O O  -  -  -
1  7 MET  CB C  -  -  40.00
1  8 MET  HB2 H  HB 1,9  8.95
1  9 MET  HB3 H  HB 1,8  8.95
1 10 MET  CG C  -  -  40.00
1 11 MET  HG2 H  HG 1,12  8.95
1 12 MET  HG3 H  HG 1,11  8.95
1 13 MET  SD S  -  -  -

```

Abbildung 10: Compound- File eines gegebenen experimentellen Spektrums. Es beschreibt für ein Compound alle vom Experiment unabhängige Daten. Insbesondere sind hier die Atombezeichnungen und die Art der Aminosäure zu finden. Meta- und Compound- File müssen über die Atom- und die Aminosäurenummer miteinander abgeglichen werden, um eine gemeinsame Datenbasis an benötigten Informationen zu erhalten.

### 3.2.2.3. Interne Datenaufbereitung

Für das Programm STEREO haben wir die eingelesenen Daten zunächst in drei Datenbasen organisiert. Eine für experimentelle Daten, eine für simulierte Daten und eine Datenbasis, in der alle Informationen, die aus Pdb, Meta- und Compound File gewonnen wurden, zusammengefasst sind.

Aus diesen drei Informationsquellen werden in einem anschließenden Schritt durch Abgleich und Verwerten von Informationen, alle Paare von Atomen oder Atomgruppen gewonnen, die im Verlauf des Programms mittels R-Wert Analyse stereospezifisch betrachtet werden müssen, d.h. es wird geprüft, wann zwei Atome oder –gruppen zueinander gehören, ob sie eindeutig definiert sind, und ob jeweils entsprechende simulierte und experimentelle Signale für diese Paare vorhanden sind.

Dabei werden den einzelnen Paaren sowohl ihre Shifts als auch ihr Volumen, der Qualitätswert und der Ambiguity- Code zugeordnet. In dieser Version von STEREO haben wir Ambiguity- Codes von 1, 2 und 4 in Betracht gezogen und entsprechend bearbeitet.

Diese Datenbasis enthält ebenfalls ein Feld, für die R- Werte im Original und im vertauschten Zustand, und ein weiteres Feld, in dem steht, ob getauscht werden soll. Außerdem wird gespeichert, welcher Aminosäure die Atome zugehören.

### 3.2.3. Programmbeschreibung und -ablauf

#### 3.2.3.1. Programmaufruf

Das Programm STEREO findet sich in AUREMOL unter dem Pull- down Menü "Calculation". Dort wird mit dem Punkt STEREO eine Eingabemaske gestartet, in der die zu verwendenden Files angegeben werden müssen. In der jetzigen Version von STEREO kann nur jeweils ein pdb- File (eine Proteinstruktur) und ein Spektrum gleichzeitig betrachtet werden. Für spätere Versionen ist allerdings geplant, Spektrenkombination und Strukturbündel zuzulassen.

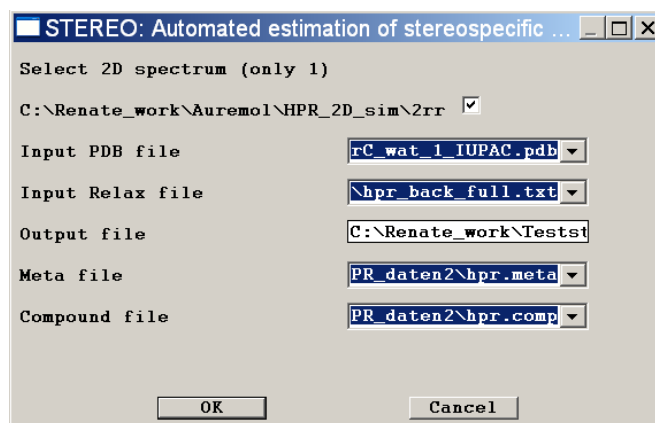


Abbildung 11: Screenshot des Anfangsdialogs von STEREO aufgerufen durch den Eintrag STEREO im Pull- down Menü „Calculation“. Dort müssen die verwendeten Filenamen angegeben werden.

Zunächst muss der Benutzer ein 2D oder 3D NOESY- Spektrum in AUREMOL auswählen. Dann wird ein rückgerechnetes (simuliertes) Spektrum benötigt (dessen output- File vorher mittels der "backcalculation"- Funktion in AUREMOL berechnet werden muss). Zusätzlich wird ein File mit der sequenziellen Zuordnung (Meta- und Compound- File) und ein File mit den strukturellen Informationen (pdb- File) benötigt.

Sind alle Eingaben korrekt, werden die angegebenen Files eingelesen, wobei Compound- und MetaFile schon zu diesem Zeitpunkt zusammengefasst werden.

### 3.2.3.2. Datenaufbereitung

Nachdem alle Daten erfasst sind, werden sie für den weiteren Verlauf des Programms so organisiert, dass ein schneller und sicherer Zugriff darauf möglich ist.

Das Feld für die experimentellen, sowie für die simulierten Daten enthält die Information über die Shifts, den Namen, die Bayssche Wahrscheinlichkeit und das Volumen der Peaks, außerdem wird der Masterliste die Angabe über Dimension und Art des Experiments entnommen.

Im Feld für zusätzliche Informationen, wie sie dem Pdb, dem Meta- und CompoundFile entnommen werden, steht der Name und die Nummer des Atoms, die chemische Verschiebung, die Art und Nummer der Aminosäure und der Ambiguity- Code.

Atomnummer	Name	AS_Nummer	AS_Art	Shift	Ambiguity
23	HA	2	LYS	3,765	1
24	HB2	2	LYS	2,345	2
25	HB3	2	LYS	2,543	2
1	HN	3	GLY	1,111	1

Abbildung 12: Beispiel für eine Datenorganisation aus Meta- und Compound- File

### 3.2.3.3. Ermitteln der stereospezifisch betrachtbaren Paare

Aus dem Pdb- File unter Abgleich mit experimentellen und simulierten Daten werden alle stereospezifisch betrachtbaren Paare ermittelt. Dazu wird zunächst entschieden, welche Atome aus dem Pdb- File Methylengruppen sind, d.h. in einer Aminosäure dürfen z.B. nur zwei  $H^\beta$  vorkommen, und diese müssen sich in der Nummerierung unterscheiden (z.B.  $H^{\beta 2}$  und  $H^{\beta 3}$ ). Dadurch wird sichergestellt, dass sowohl Daten im IUPAC- Format, als auch im CNS- Format verarbeitet werden können. (Gleiches gilt für  $H^{\delta,\gamma}$ , etc.). Auch werden Methylengruppen an Seitenketten- Ringen betrachtet.

Aus dieser ersten Abfrage resultiert ein Datenfeld mit allen überhaupt denkbaren Paaren. Bei der Generierung dieses Feldes werden gleichzeitig auch schon Spalten für die Berechnung des R- Wertes und ihrem Vorhandensein in sowohl der experimentellen als auch simulierten Masterliste angelegt, die später gefüllt werden.

In einem nächsten Schritt wird der Ambiguity- Code der Eingabedaten überprüft. Wird ein Code von 1 (d.h. dieser Shift wurde manuell stereospezifisch exakt zugeordnet und darf nicht verändert werden) ermittelt, wird dieses Paar von der weiteren Betrachtung ausgeschlossen.

Name	Name	AS_Nummer	Shift	Shift	Ambiguity	in exp vorhanden	in calc vorhanden	rWert original	rWert getauscht	tauschen
HB2	HB3	5	2,345	2,54	2					
HB2	HB3	6	1,234	1,68	2					
HD11	HD12	7	5,123	4,99	2					
HB2	HB3	11	1,789	2	2					
HG1	HG2	11	3,444	3,12	1					

*Abbildung 13: Erster Schritt der Aufbereitung der stereospezifisch betrachtbaren Paare. Alle möglichen Paare werden aus dem pdb- File ermittelt, diejenigen mit Ambiguity- Code 1 ausgeschlossen.*

Anschließend werden die ermittelten stereospezifischen Paare im experimentellen und simulierten Spektrum gesucht, denn nur, wenn sie in beiden Files vorhanden sind, können sie später auch mittels R-Wert Vergleich überprüft werden.

Ist eines dieser Paare, oder auch nur ein Atom nicht in allen beiden Dateien vorhanden, wird dieses Paar nicht weiter betrachtet.

Name	Name	AS_Nummer	Shift	Shift	Ambiguity	in exp vorhanden	in calc vorhanden	rWert original	rWert getauscht	tauschen
HB2	HB3	5	2,345	2,54	2	ja	ja			
HB2	HB3	6	1,234	1,68	2	nein	ja			
HD11	HD12	7	5,123	4,99	2	ja	ja			
HB2	HB3	11	1,789	2	2	ja	ja			

*Abbildung 14: Zweiter Schritt der Aufbereitung der stereospezifisch betrachtbaren Paare. Paare, die nicht in allen beiden Dateien vorkommen, werden ausgeschlossen.*

Ein zusätzlicher Sonderfall entsteht hier bei der Betrachtung von Atomen mit Ambiguity- Code 4. Dies repräsentiert Atomgruppen einer Aminosäure, deren Resonanzen nicht genau



ermittelt werden können. So z.B. die  $H_\gamma$  und die  $H_\delta$  einer Aminosäure. Dabei weiß man bei zwei im Spektrum gegebenen Signalen zwar, dass eines von beiden zu den  $H_\gamma$ , das andere zu den  $H_\delta$  gehört, aber nicht, welches zu welchem.

Diese Paare werden in einem Zwischenschritt, in dem u.a. überprüft wird, ob in einer Aminosäure auch nur genau zwei solcher Signale vorhanden sind, aus den gegebenen Informationen herausgefiltert, und der Liste der zu überprüfenden Paare hinzugefügt.

Wurde auf diese Weise die Liste der zu überprüfenden stereospezifischen Zuordnungen reduziert, verifiziert, vervollständigt und optimal an die Daten angepasst, kann im nächsten Schritt die Berechnung der unterschiedlichen R- Werte erfolgen.

### 3.2.3.4. Berechnen der R- Werte

Bei der Überprüfung der stereospezifischen Zuordnung werden alle in der Liste der Paare vorhandenen Einträge nacheinander abgearbeitet.

Von jedem Paar wird zunächst ein lokaler R- Wert in der Originalzuordnung berechnet. Dieser Wert wird im entsprechenden Feld gespeichert.

Die in diesem Fall verwendete Formel zur Berechnung des R- Wertes erfolgt nach Gleichung 6. Es werden jeweils alle Signale verwendet, durch die das betrachtete Atompaar mit allen anderen Atomen in Beziehung steht.

Danach wird jedes Paar im experimentellen File gesucht, und alle Benennungen werden vertauscht. Anschließend wird wieder eine R- Wert Berechnung durchgeführt. Auch dieser Wert wird gespeichert. Hat sich der Wert verbessert, wird diese Atom- Kombination als „zu tauschend“ klassifiziert.

Name	Name	AS_Nummer	Shift	Shift	Ambiguity	in exp vorhanden	in calc vorhanden	rWert original	rWert getauscht	tauschen
HB2	HB3	5	2,345	2,54	2	ja	ja	0,076	0,023	ja
HD11	HD12	7	5,123	4,99	2	ja	ja	0,121	0,222	nein
HB2	HB3	11	1,789	2	2	ja	ja	0,045	0,012	ja

*Abbildung 15: Endzustand der Informations- Tabelle. Alle benötigten Werte zur Optimierung der stereospezifischen Zuordnung sind vorhanden.*

Für die Berechnung der R- Werte haben wir eine Routine eingebaut, die einen lokalen R- Wert einer gegebenen Aminosäure berechnet und zurückgibt. Zusätzlich musste ein Skalierungsfaktor (vgl. Gleichung 6) ermittelt werden, der die Volumina der simulierten Signale auf die der experimentellen abbildet, da diese sich durch unterschiedliche Messbedingungen, wie z.B. Proteinkonzentration, Spektrometer oder Probenkopf, stark voneinander unterscheiden.

Beide Routinen sind analog der entsprechenden Routinen im Programm RFAC, das in AUREMOL integriert ist, und bereits beschrieben wurde [Gro00].

Durch die Integration dieser Unterprogramme in STEREO wurde es allerdings möglich, den Ablauf des Programms stark zu beschleunigen.

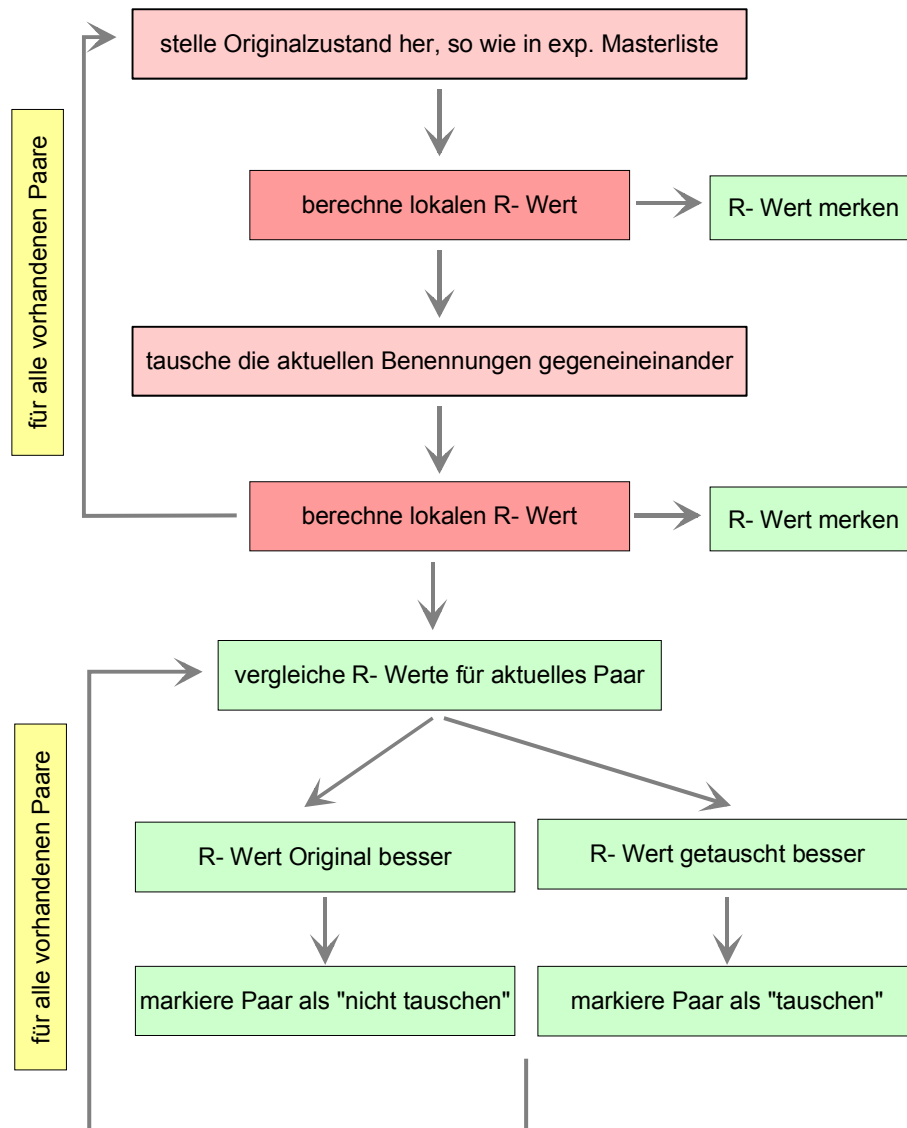


Abbildung 16: Schematischer Ablauf der Berechnung der unterschiedlichen R-Werte und anschließender Vergleich.

Wie sich im Verlauf der Testreihen (vgl. 3.3.2.2 Testergebnisse) herausgestellt hat, ist es auch nötig, bei der Berechnung der R- Werte zu überprüfen, ob ein spezielles Signal, das zur Berechnung der R- Faktoren herangezogen wird, sowohl im experimentellen, als auch im rückgerechneten File vorhanden ist. Ist das nicht der Fall, kann es in Einzelfällen vorkommen, dass sich der lokale R- Wert dadurch verbessert, dass nach dem Tausch das betreffende Signal nicht mehr vorhanden ist, und somit auch nicht zur Berechnung beiträgt. Dieser Fall entsteht, wenn z.B. ein  $H^{\beta 2}$ - Atom von einem anderen Atom 4 Angström entfernt ist, wogegen das  $H^{\beta 3}$ - Atom zu diesem Atom 6 Angström Distanz aufweist. In der Simulation, die ab einem vorgegebenen Distanzkriterium (normalerweise 5 Angström) keine Signale mehr zurückrechnet, ist deshalb das zweite Signal nicht mehr vorhanden. Abbildung 17 und Abbildung 18 veranschaulichen die Situation.

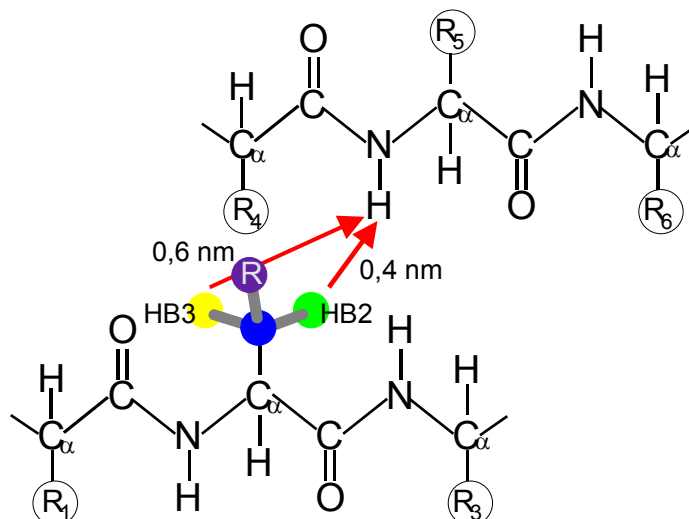


Abbildung 17: Darstellung eines Spezialfalles, bei dem in der Rückrechnung nur das Signal von  $H^{\beta 2}$  zum  $H^N$  simuliert wird, wogegen das Signal von  $H^{\beta 3}$  zu diesem  $H^N$  fehlt.

Bei einem Tausch der Atompaarbenennung in der experimentellen Masterliste wird somit dieses betreffende Signal bei der R- Wert- Berechnung zu den unzugeordneten Signalen hinzugefügt.

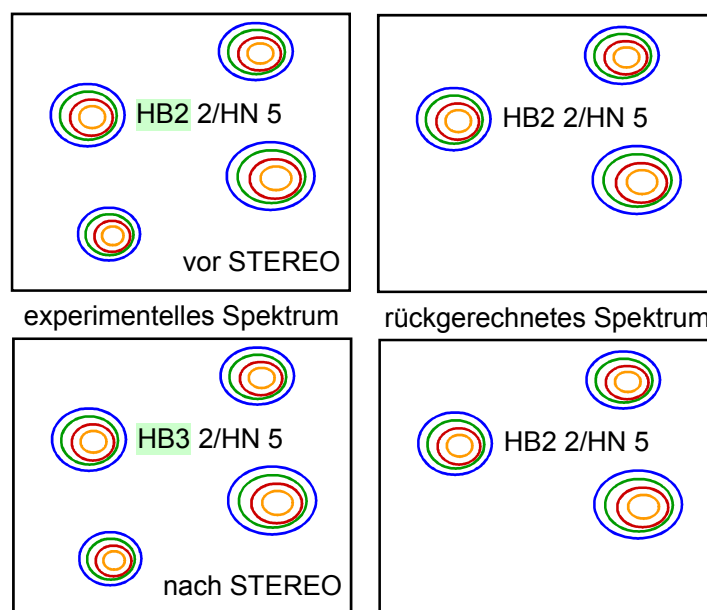
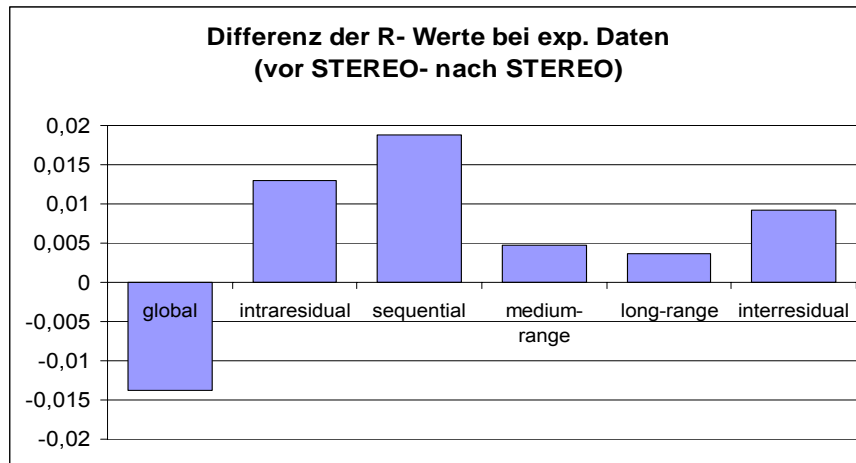


Abbildung 18: Auswirkungen eines von STEREO durchgeführten Tausches im Spezialfall von Abbildung 17. Oben dargestellt die Situation vor STEREO: für das Signal  $H^{\beta 2}$  2 zu  $H^N$  5 ist in der Rückrechnung ein korrespondierender Peak vorhanden. Unten die Situation nach einem Tausch der Benennung der Atompaare: Für das Signal  $H^{\beta 3}$  2 zu  $H^N$  5 existiert kein korrespondierender Peak, dieses Signal wird also der Menge der unzugeordneten Signale zugerechnet.

Da der globale R- Wert (berechnet aus allen long- Range- Signalen unter Berücksichtigung der unzugeordneten Signale, vgl. [Gro00]) sehr stark von vorhandenen unzugeordneten Signalen abhängt, ist es in diesen Fällen, da sich ja die Anzahl der unzugeordneten Signale erhöht hat, unvermeidlich dass dieser sich verschlechtert, obwohl sich die Zuordnung (und damit alle R- Faktoren, die ohne unzugeordnete Signale berechnet wurden) verbessert hat. Abbildung 19 illustriert diesen Fall.



*Abbildung 19: Übersicht über die unterschiedlichen R- Werte. Deutlich zu sehen ist die Verschlechterung des globalen R- Wertes, der daher rührt, dass nach dem Lauf von STEREO ca. 5% mehr unzugeordnete Signale vorhanden sind, als vorher.*

Durch den Einbau oben genannter Abfrage konnte dieser Effekt beseitigt werden.

Natürlich könnte dieser Effekt auch dadurch beseitigt werden, dass in der Rückrechnung kein Abschneidekriterium eingefügt, und alle nur denkbaren Signale rückgerechnet werden. Aufgrund der dabei aber anfallenden Datenmengen und der unverhältnismäßig langen Rechenzeit ist das für die Normalanwendung kein praktikabler Lösungsweg.

Da es nun jedoch bei experimentellen Spektren außerdem immer unzugeordnete Signale gibt, die sich durch die neue stereospezifische Zuordnung auch verändern, worauf, wie bereits erwähnt, der globale R- Wert sehr sensibel reagiert, haben wir für die Überprüfung der stereospezifischen Zuordnung einen neuen R- Faktor eingeführt, der sich aus allen Signalen (intra- und interresiduell) ohne die unzugeordneten berechnet<sup>2</sup>. Dieser R- Faktor ist unabhängig von unzugeordneten Signalen und spiegelt daher optimal die kleinen Veränderungen in der räumlichen Struktur wieder, die durch eine geänderte stereospezifische Zuordnung entstehen.

<sup>2</sup> Zum Vergleich: der globale R- Wert berechnet sich aus allen long- range ( $i \geq 4$ ) Signalen unter Einbeziehung der unzugeordneten Peaks.

### 3.2.3.5. Generieren der Ausgabedaten

Sind alle Berechnungen durchgeführt, wird anhand der Einträge in der Informations-Datenbasis entschieden, welche Benennungen besser getauscht werden sollten, und welche so bleiben, wie sie sind.

Wurde ein Paar als „zu tauschend“ klassifiziert, wird intern im experimentellen Datensatz die Benennung von z.B.  $H^{\beta 2}$  und  $H^{\beta 3}$  einer Aminosäure geändert.

Diese Änderungen werden bei allen betroffenen Paaren durchgeführt, danach werden die neuen Ergebnisse einerseits in die originale Masterliste übernommen, andererseits wird ein zusätzliches Ausgabe- File erzeugt, in dem die errungenen Erkenntnisse in „plain Text“ festgehalten werden.

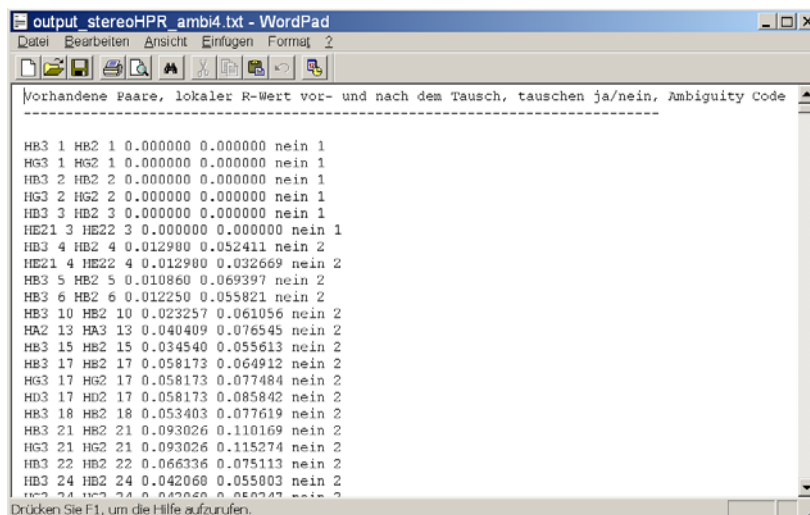
## 3.3. Ergebnisse

### 3.3.1. Programmergebnis

Wenn das Programm beendet wurde, informiert ein Fenster den Benutzer darüber, ob STEREO erfolgreich beendet wurde.

Die Masterliste wird automatisch aktualisiert, und der Benutzer kann das Ergebnis direkt am Bildschirm anhand der benannten Signale betrachten.

Zusätzlich zur Aktualisierung der Masterliste nach dem Programmlauf, wird ein textbasiertes Ausgabe- File erzeugt, in dem alle Paare, die beim ersten Schritt gefunden wurden aufgelistet sind, mit allen dazugehörenden Informationen, wie R- Werte, und ob getauscht wurde. Außerdem werden die Daten der geänderten Masterliste zusätzlich im „plain- Text“ Format ausgegeben, was eine Weiterverarbeitung in anderen Programmen, z.B. EXCEL möglich macht.

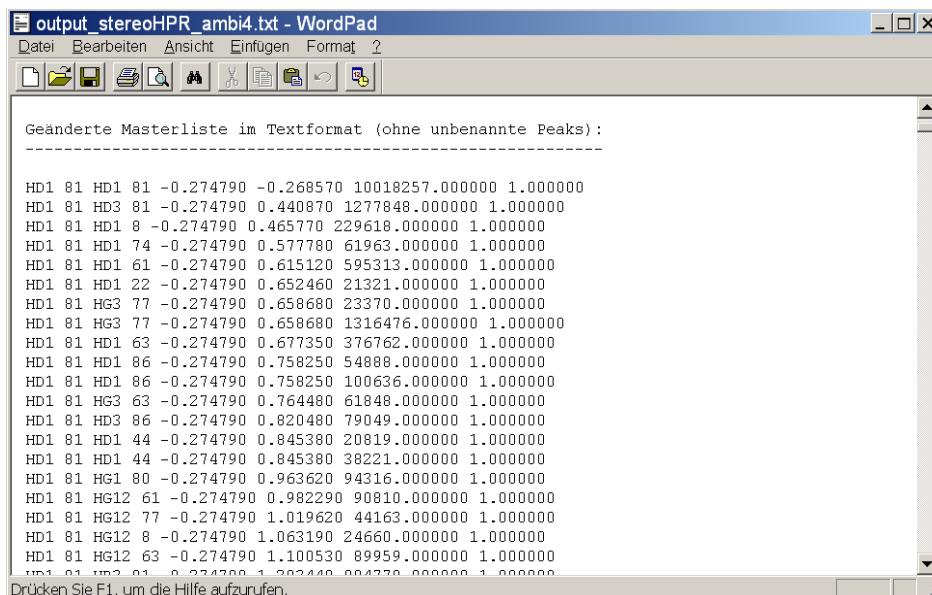


```

Vorhandene Paare, lokaler R-Wert vor- und nach dem Tausch, tauschen ja/nein, Ambiguity Code
-----
HB3 1 HB2 1 0.000000 0.000000 nein 1
HG3 1 HG2 1 0.000000 0.000000 nein 1
HB3 2 HB2 2 0.000000 0.000000 nein 1
HG3 2 HG2 2 0.000000 0.000000 nein 1
HB3 3 HB2 3 0.000000 0.000000 nein 1
HE21 3 HE22 3 0.000000 0.000000 nein 1
HB3 4 HB2 4 0.012900 0.052411 nein 2
HE21 4 HE22 4 0.012900 0.032669 nein 2
HB3 5 HB2 5 0.010860 0.069397 nein 2
HB3 6 HB2 6 0.012250 0.055821 nein 2
HB3 10 HB2 10 0.023257 0.061056 nein 2
HA2 13 HA3 13 0.040409 0.076545 nein 2
HB3 15 HB2 15 0.034540 0.055613 nein 2
HB3 17 HB2 17 0.058173 0.064912 nein 2
HG3 17 HG2 17 0.058173 0.077484 nein 2
HD3 17 HD2 17 0.058173 0.085842 nein 2
HB3 18 HB2 18 0.053403 0.077619 nein 2
HB3 21 HB2 21 0.093026 0.110169 nein 2
HG3 21 HG2 21 0.093026 0.115274 nein 2
HB3 22 HB2 22 0.066336 0.075113 nein 2
HB3 24 HB2 24 0.042068 0.055803 nein 2
HB3 24 HB2 24 0.042068 0.055803 nein 2

```

Abbildung 20: Output- File von STEREO. Alle im Pdb- File ermittelten Paare werden aufgelistet, die ermittelten R- Werte im Originalzustand und bei vertauschter Benennung, zusätzlich eine Angabe darüber, ob getauscht wurde oder nicht. In der letzten Spalte stehen die Ambiguity- Codes der einzelnen Paare. Wurde ein Paar aufgrund eines Ambiguity- Codes von 1 nicht behandelt, wird ein R- Wert von 0 ausgegeben.



```

Geänderte Masterliste im Textformat (ohne unbenannte Peaks):
-----
HD1 81 HD1 81 -0.274790 -0.268570 10018257.000000 1.000000
HD1 81 HD3 81 -0.274790 0.440870 1277848.000000 1.000000
HD1 81 HD1 8 -0.274790 0.465770 229618.000000 1.000000
HD1 81 HD1 74 -0.274790 0.577780 61963.000000 1.000000
HD1 81 HD1 61 -0.274790 0.615120 595313.000000 1.000000
HD1 81 HD1 22 -0.274790 0.652460 21321.000000 1.000000
HD1 81 HG3 77 -0.274790 0.658680 23370.000000 1.000000
HD1 81 HG3 77 -0.274790 0.658680 1316476.000000 1.000000
HD1 81 HD1 63 -0.274790 0.677350 376762.000000 1.000000
HD1 81 HD1 86 -0.274790 0.758250 54888.000000 1.000000
HD1 81 HD1 86 -0.274790 0.758250 100636.000000 1.000000
HD1 81 HG3 63 -0.274790 0.764480 61848.000000 1.000000
HD1 81 HD3 86 -0.274790 0.820480 79049.000000 1.000000
HD1 81 HD1 44 -0.274790 0.845380 20819.000000 1.000000
HD1 81 HD1 44 -0.274790 0.845380 38221.000000 1.000000
HD1 81 HD1 80 -0.274790 0.963620 94316.000000 1.000000
HD1 81 HG12 61 -0.274790 0.982290 90810.000000 1.000000
HD1 81 HG12 77 -0.274790 1.019620 44163.000000 1.000000
HD1 81 HG12 8 -0.274790 1.063190 24660.000000 1.000000
HD1 81 HG12 63 -0.274790 1.100530 89959.000000 1.000000
HD1 81 HD2 81 -0.274790 1.202440 884738.000000 1.000000

```

Abbildung 21: Output- File von STEREO. Im gleichen File werden auch die benannten experimentellen Daten der Masterliste ausgegeben, unabhängig davon, ob sie von STEREO behandelt wurden, oder nicht. Aufgelistet ist die Benennung der Signale, ihre chemische Verschiebung, das Volumen und der Qualitätswert der Peaks.

### 3.3.2. Tests und Anwendungsbeispiele

#### 3.3.2.1. Verwendete Strukturen

Für die Tests am Programm STEREO haben wir 2D  $^1\text{H}$ - NOESY- Spektren von HPr, *TmCsp* und RalGDS benutzt, die mit 800 MHz aufgenommen wurden.

HPr, das Histidin enthaltende Protein aus *Staphylococcus carnosus*, besteht aus 88 Aminosäureresten und die Struktur stellt sich aus drei  $\alpha$ - Helices und einem viersträngigen  $\beta$ - Faltblatt dar [Goe99b].

*TmCsp* ist ein Kälteschock- Protein aus *Thermotoga maritima* und besteht aus 66 Aminosäureresten, die ein fünfsträngiges  $\beta$ - Faltblatt und eine  $\alpha$ - Helix bilden [Har98].

Die Struktur des RalGDS- Proteins (Ras- Bindedomäne des Ral- Guanosin- Dissoziationsstimulators aus *Homo sapiens*) besteht aus 93 Aminosäureresten und stellt sich aus zwei  $\alpha$ - Helices und einem fünfsträngigen  $\beta$ - Faltblatt dar. Diese Struktur stammt aus der Bio- Mag- Res- Bank [Bmr01], und wurde 1999 von T.D.Müller, L.Handel, P.Schmiederer, und H.Oschkinat publiziert.

Für HPr haben wir aus der o.g. Ausgangsstruktur vier weitere Test- Strukturen erzeugt. Dazu wurden 3000, 6000 und 9000 Schritte von 0,005 ps MD- Läufen im Vakuum bei Raumtemperatur simuliert, um eine schrittweise ansteigende Unordnung in der Struktur zu erreichen. Die letzte Struktur war ein energieminimierter ausgestreckter Strang.

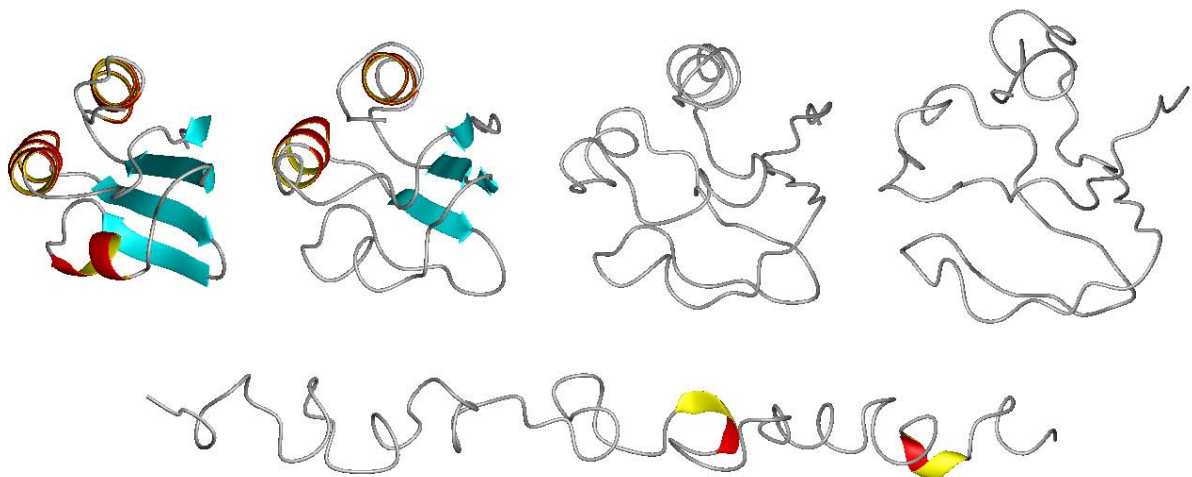


Abbildung 22: Die verschiedenen Strukturen des HPr. Ganz links die Ausgangsstruktur, darauf folgend die Strukturen nach 3000, 6000 und 9000 MD- Läufen im Vakuum. Unten der energieminimierte ausgestreckte Strang.

Außerdem wurde für HPr zusätzlich ein Bündel aus 13 endgültigen NMR- Strukturen für weitere Untersuchungen verwendet.

### 3.3.2.2. Testergebnisse

In der neuen Version wurde STEREO von uns zunächst mit synthetischen Daten getestet. Dazu fanden simulierte Spektren von *TmCSP* und HPr Verwendung. Die Simulation wurde mit einer cutoff distance von 0,6nm, einer Mischzeit von 0,1s, bei einer Frequenz von 600 Mhz durchgeführt. Es wurden in der w1 Dimension 1024, und in w2 2048 Datenpunkte berechnet. Außerdem wurde eine J- Korrektur für die Mischzeit  $T_2$  durchgeführt (vgl. [Goe94]).

Als erstes haben wir simulierte Masterlisten erstellt, und diese dann mittels STEREO auf ihre stereospezifische Zuordnung überprüft. Wie zu erwarten war, wurde bei *TmCSP* und HPr in 100% der Fällen die Aussage getroffen, dass keine der Benennung falsch ist, und es wurde nichts getauscht.

Daraufhin wurden alle stereospezifisch in Frage kommenden Paare ( $H^\beta$ ,  $H^\gamma$ ,  $H^\delta$ ,  $H^\epsilon$ ) manuell vertauscht, und danach wieder mit STEREO überprüft. Diesmal wurde bei *TmCSP* durch den Programmlauf in allen Fällen die Benennung geändert, weil sich im vertauschten Zustand ein besserer R- Wert ergab. Bei HPr wurden alle bis auf ein Signal vertauscht. Dieses eine Signal wies den Sonderfall eines gleichen R- Wertes sowohl im Original, als auch im vertauschten Zustand auf. In diesen Fällen wird die ursprüngliche Benennung belassen.

CSP sim		HPr sim	
Paare	82	Paare	83
tauschen	0	tauschen	0
nach Vertauschen		nach Vertauschen	
tauschen	82	tauschen	82
Ausgangsposition	nach Vertauschen	Ausgangsposition	nach Vertauschen
0,0%	100,0%	0,0%	98,8%

Abbildung 23: Ergebnis des Tests mit simulierten Daten. Dargestellt ist die Anzahl der vertauschten Paare bei unveränderten simulierten Daten und bei manuell vertauschten Daten.

Nachdem durch diesen einfachen Test die richtige Funktion von STEREO sichergestellt war, haben wir anhand simulierter Daten reelle Bedingungen emuliert. Dazu wurde die originale 3D Struktur von HPr mit der dazugehörenden Zuordnung verwendet um einerseits einen perfekten Datensatz mit dazu genau passendem rückgerechnetem Spektrum und andererseits durch Moleküldynamik- Läufe, bei denen sich die gefaltete Struktur schrittweise entfaltet, veränderte Daten zu erhalten.



Die Ausgangsstruktur wurde durch 3000, 6000 und 9000 Moleküldynamik- Läufe aufgeweitet, daraus wurde dann jeweils wieder ein Spektrum rückgerechnet, was einem experimentellen Spektrum mit unvollständiger und fehlerhafter Zuordnung entspricht.

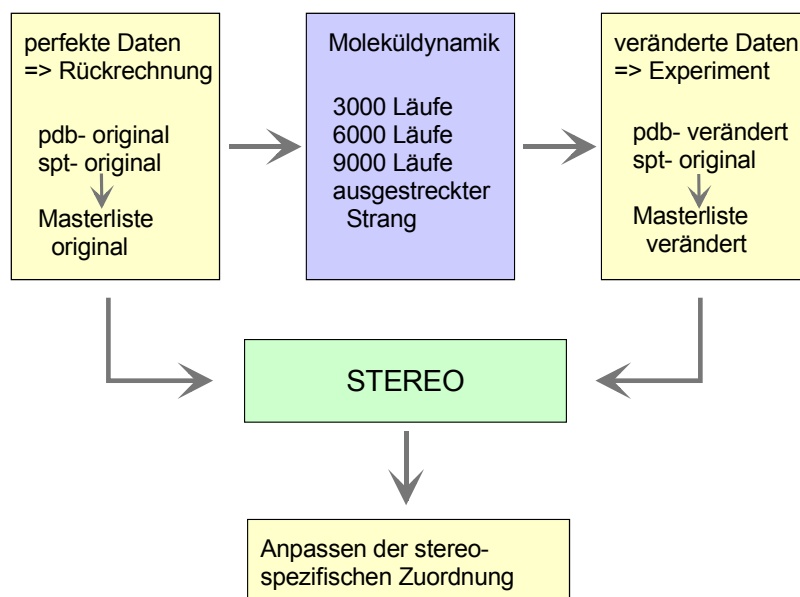


Abbildung 24: Schema des Tests mit schrittweise verschlechterten simulierten Daten.

Die Ergebnisse (vgl. Abbildung 25) zeigen zunächst wie erwartet in der Ausgangsstruktur keine falsch zugeordneten stereospezifischen Signale, was logisch ist, da das Spektrum aus der Struktur generiert wurde. Bringt man dann die Struktur schrittweise in Unordnung, so nimmt auch die Zahl der stereospezifisch jetzt zwangsläufig falsch zugeordneten Signale zu, da nun Struktur und Zuordnung nicht mehr ideal zusammenpassen.

Nun könnte man vermuten, dass mit steigender Unordnung auch mehr stereospezifische Signale falsch sind und daher immer mehr getauscht werden müssen. Dabei ist es aber statistisch so, dass bei einer zufälligen Verteilung der Methyl- und Methylengruppen eine Tauschrate von etwas unter 50% zu erwarten ist, wenn eine „genügende“ Unordnung erreicht ist, was auch in Abbildung 25 zu sehen ist<sup>3</sup>. Natürlich ist es so, dass in Einzelfällen etwas mehr oder weniger als diese zu erwartende Hälfte auftreten wird. Generell kann man sehen, dass wohl in den Testdaten schon nach dem ersten, spätestens aber nach dem zweiten MD- Lauf die Gruppen nur mehr zufälligen Anordnungen folgen.

<sup>3</sup> etwas weniger als 50% deshalb, da in manchen Aminosäuren die stereospezifischen Gruppen aufgrund der Geometrie dieser Aminosäure festgelegt sind, und deshalb auch keine andere Konformation einnehmen können.

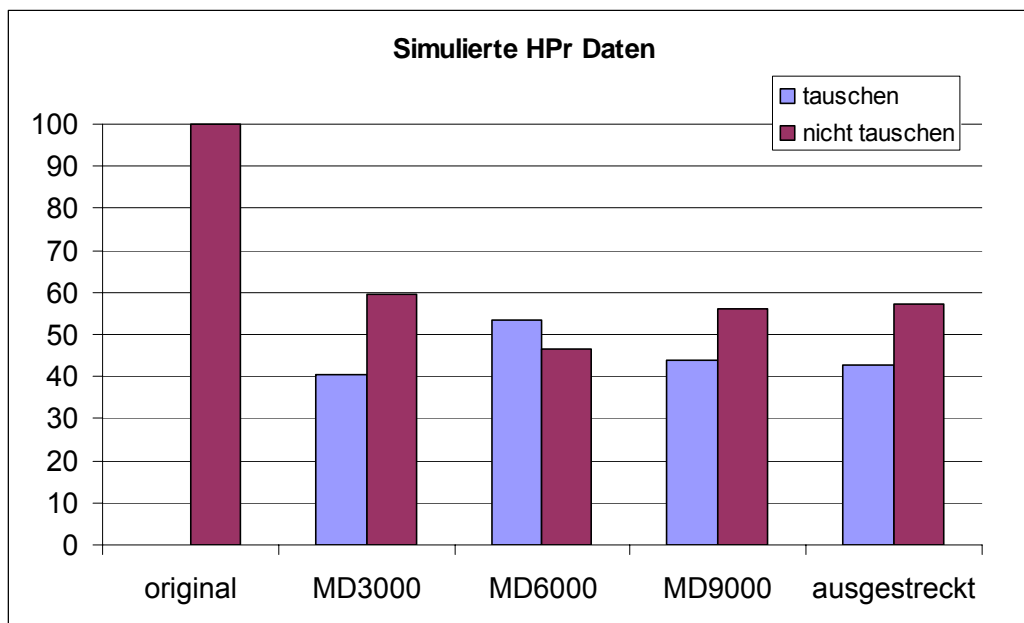


Abbildung 25: Anzahl der zu tauschenden Paare in den jeweiligen Zuordnungen für die angegebenen MD- Läufe in %.

Um generelle und gesicherte Unterschiede in der Verteilung der zu tauschenden Signale bei zunehmender Unordnung sichtbar zu machen, müssten für mehrere hundert Strukturen verschiedene Moleküldynamik- Schrittzahlen berechnet werden. Danach müssten die Strukturen dann jeweils auf ihre stereospezifischen Signale hin untersucht werden.

Da es bei diesem Test jedoch lediglich auf die richtige Funktionsweise des Programms ankam, wurde auf diese zeitraubende Statistik verzichtet.

Auch für die Behandlung der unterschiedlichen Ambiguity- Codes (vgl. Abschnitt 3.2.3.3) haben wir Tests durchgeführt, die zeigten, dass STEREO korrekt arbeitet.

Zum Schluss der kurzen Testreihe soll zusammenfassend ein kleiner Blick auf die Verteilung der diastereotopen Atome im Vergleich zur Gesamtzahl der H- Atome geworfen werden.

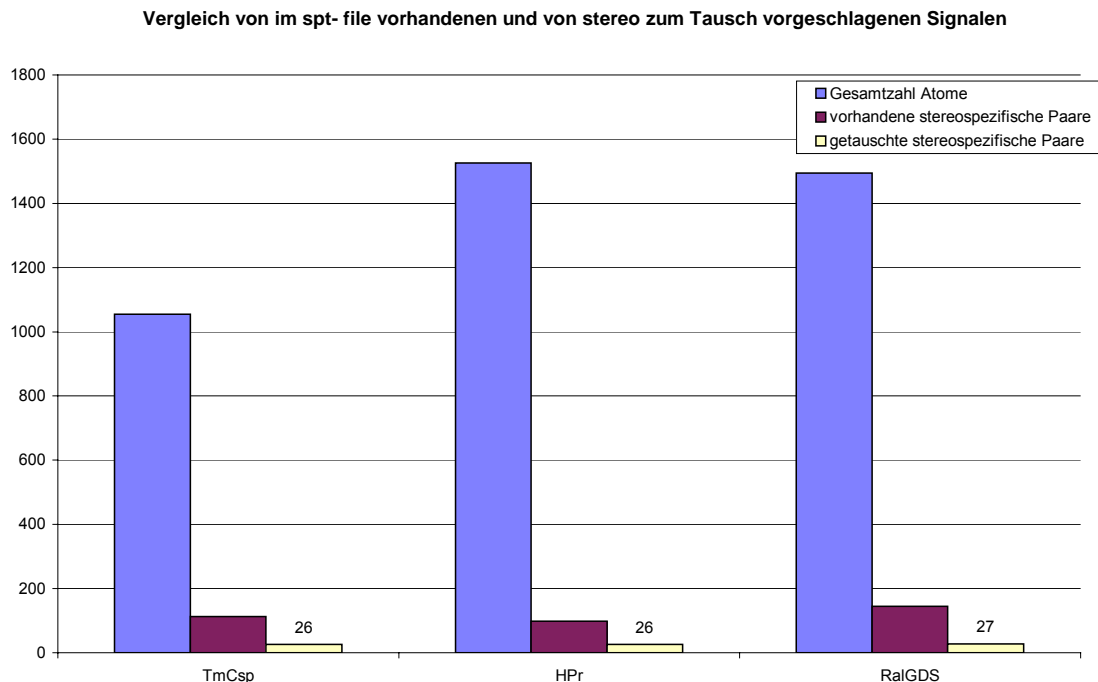
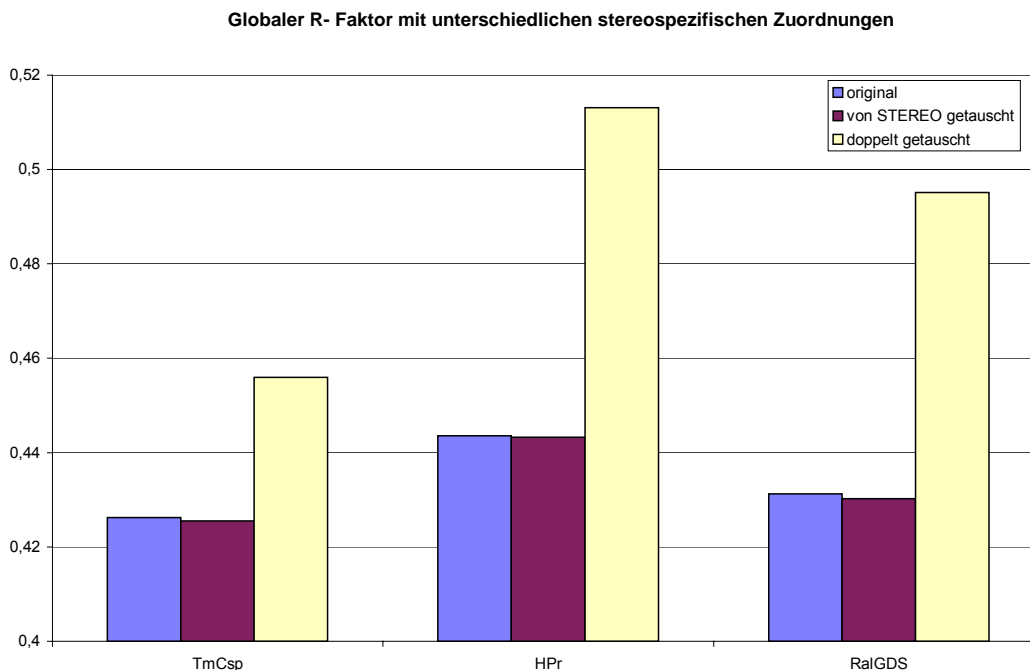


Abbildung 26: Geänderte stereospezifische Zuordnungen in den Zuordnungslisten. Überblick über die absolute Zahl vorhandener H- Atome im Vergleich zu den stereospezifischen Signalen und den getauschten Signalen. Die relativ geringe Anzahl führt zu einer vergleichsweise geringen Verbesserung des globalen R- Wertes.

Wenn man die Verteilung der stereospezifischen Atome in den Zuordnungslisten betrachtet, sieht man, dass nur bei einem geringer Teil der stereospezifischen Atome die Benennung geändert wird, und dass die stereospezifischen Atome in ihrer Gesamtheit wiederum nur einen geringen Teil im Verhältnis zu allen vorhandenen Atomen ausmachen.

Das wiederum erklärt den nur kleinen Effekt der Verbesserung, den die richtige stereospezifische Zuordnung auf den globalen R- Wert zeigt, wie schon an anderer Stelle ausgeführt worden ist [Kir01].

Abbildung 27 zeigt unterschiedliche globale R- Werte für unterschiedliche stereospezifische Zuordnungen manuell zugeordneter experimenteller Spektren der drei Testproteine TmCsp, HPr und RalGDS. Das zweimalige Tauschen der Zuordnung (tauschen der Protonenpaare nach der richtigen Zuordnung durch STEREO) bewirkt hier, dass alle vorkommenden stereospezifischen Paare falsch benannt sind. Wie man sieht, ist der R- Wert am geringsten bei richtiger Zuordnung, und signifikant höher, wenn alle stereospezifischen Signale falsch benannt sind.



*Abbildung 27: Globaler R- Faktor (long- range Signale mit unzugeordneten, vgl. [Gro04]) mit unterschiedlichen stereospezifischen Zuordnungen. Der globale R- Faktor wurde vor der Benutzung von STEREO, nach der erneuten stereospezifischen Zuordnung durch STEREO und nachdem das Assignment zweimal getauscht wurde, berechnet.*

In einer weiteren Testreihe, in der der Einfluss der richtigen stereospezifischen Zuordnung auf das Verhalten der einzelnen R- Werte genauer untersucht werden soll, finden drei Datensätze als „experimentelle Daten“ Verwendung, die ausgehend von einem experimentell gemessenen 2D NOESY- Spektrum von *HPr* bei 800 MHz, der dazugehörigen sequenziellen Zuordnung und der daraus resultierenden 3D- Struktur erzeugt werden.

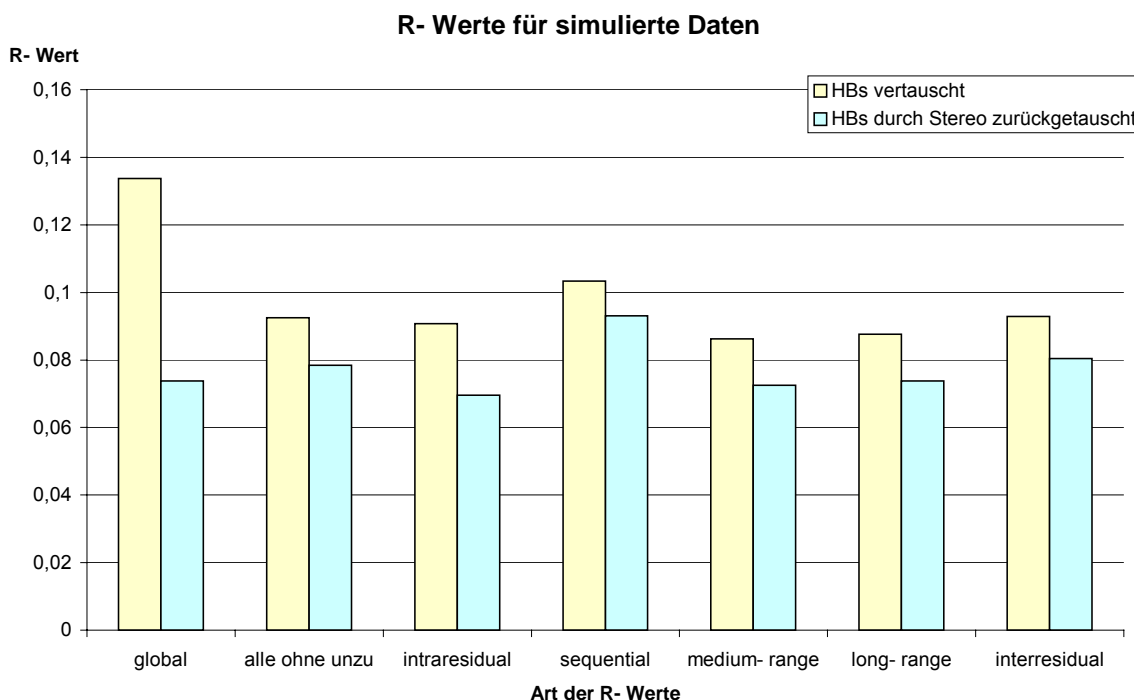
Dazu wird zunächst aus der sequenziellen Zuordnung und der Struktur ein simuliertes Spektrum generiert, das im ersten Fall ohne Veränderung verwendet, und im zweiten Fall neu gepeakpickt und automatisch zugeordnet wird.

Der dritte Datensatz besteht aus der automatischen Zuordnung (vgl. Abschnitt 1) des experimentellen Spektrums ausgehend von der sequenziellen Zuordnung und der vorhandenen Struktur.

Die Auswertung und Überprüfung dieser Daten, speziell der experimentellen, hat es erforderlich gemacht, einerseits leichte Modifikationen am Programm STEREO durchzuführen, und andererseits einen zusätzlichen, für die Betrachtung dieser relativ geringen Effekte gut geeigneten R- Wert zu definieren. Dieses wurde bereits weiter oben beschrieben (3.2.3.4).

### **R- Werte für simulierte Daten**

Im Test mit simulierten Daten wurden alle in der Masterliste vorhandenen Signale der H $\beta$ -Atome aller Aminosäuren vertauscht. Was bedeutet, dass nun alle H $\beta$ - Signale falsch benannt waren. Danach wurden R- Werte für diesen Fall bestimmt, und STEREO ausgeführt. Das Programm war in der Lage 42 von 44 vorhandenen Vertauschungen rückgängig zu machen. Auch danach wurde wieder eine R- Wert- Berechnung durchgeführt. Abbildung 28 zeigt das Ergebnis.



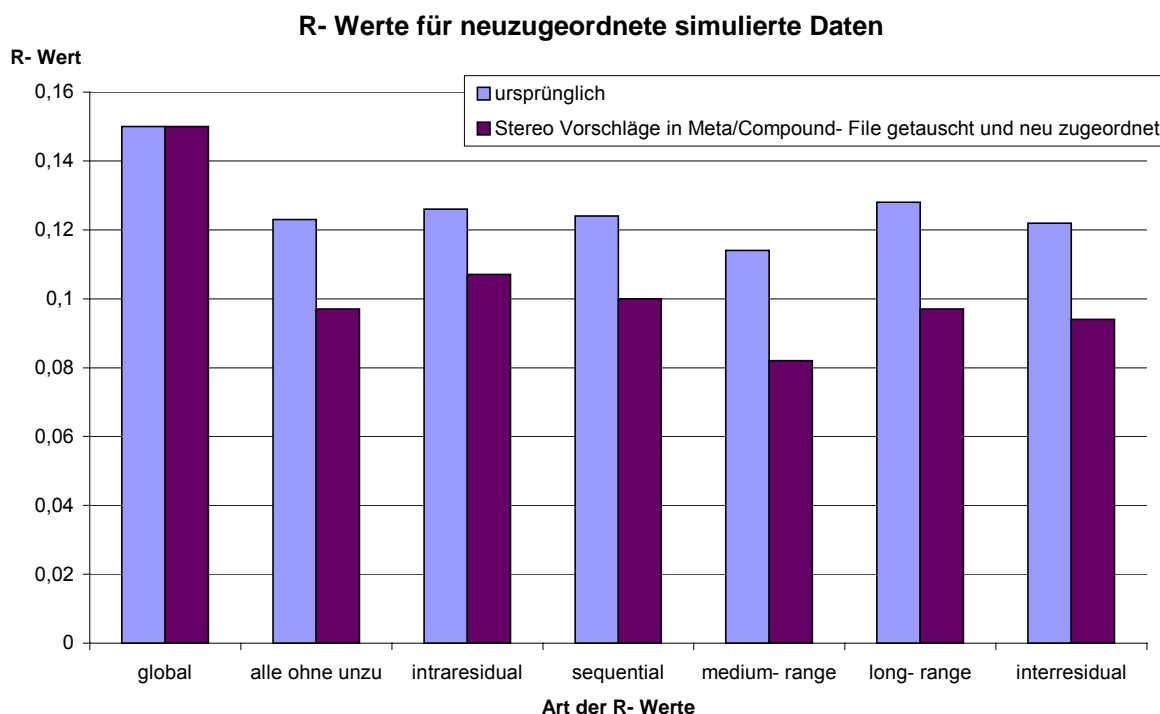
*Abbildung 28: Unterschiedliche R- Werte bei simulierten Daten von HPr. Die gelben Balken zeigen R- Werte bei falsch benannten H $\beta$ - Signalen, die grünen Balken die R- Werte nach der Richtigstellung der Zuordnung durch STEREO.*

Man sieht für alle betrachteten R- Werte eine Verbesserung nach dem Ausführen des Programms und der Richtigstellung der stereospezifischen Zuordnung durch STEREO. Insgesamt sind alle, insbesondere der globale R- Wert sehr gering, was durch das Fehlen unzugeordneter Signale begründet ist. Deshalb ist in diesem Fall auch der Effekt des Tausches auf den globalen R- Wert ausgeprägt und deutlich sichtbar.

### **R- Werte für neu zugeordnete simulierte Daten**

Im zweiten Test haben wir das simulierte Spektrum automatisch neu zugeordnet, danach wurden die R- Werte berechnet, und anschließend die Zuordnung mittels STEREO überprüft. Die von STEREO zum Tausch vorgeschlagenen stereospezifischen Benennungen wurden anschließend in den Zuordnungslisten (Meta-/Compoundfile) korrigiert, und mit diesen Daten eine erneute automatische Zuordnung durchgeführt.

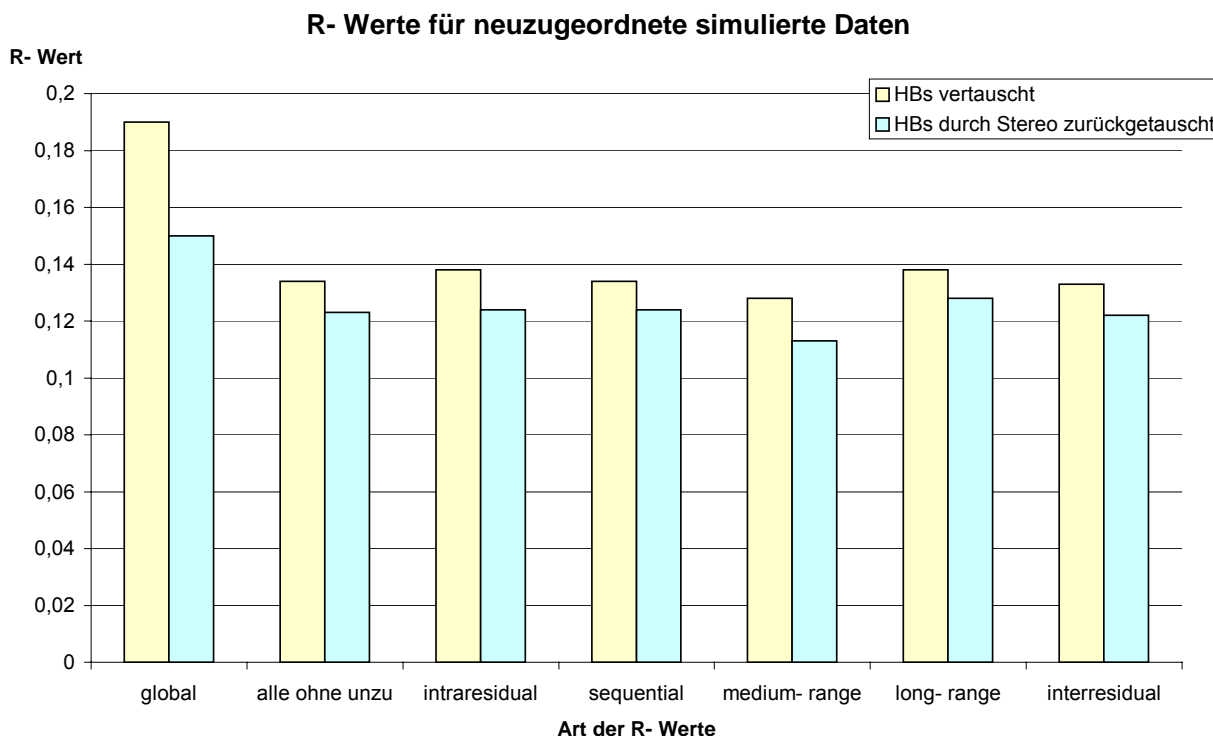
Die R- Faktoren nach dieser erneuten automatischen Zuordnung wiesen signifikant bessere Werte auf (siehe Abbildung 29).



*Abbildung 29: Unterschiedliche R- Werte bei neuzugeordneten simulierten Daten von HPr. Die blauen Balken zeigen R- Werte bei ursprünglicher Zuordnung, die violetten Balken die R- Werte nach der Richtigstellung der Zuordnung durch STEREO und einer danach erfolgten automatischen Neuzuordnung.*

In diesem Fall ist zu erkennen, dass der Effekt der richtigen stereospezifischen Zuordnung auf den globalen R- Wert keinen Einfluss hat, was einerseits auf die relativ geringe Anzahl zu tauschender Atombenennungen (8 von 71 vorhandenen), andererseits auf das Vorhandensein unzugeordneter Signale zurückzuführen ist. Der zusätzlich definierte R- Wert aus allen Signalen außer den unzugeordneten (zweites Balkenpaar von links) dagegen zeigt eine deutliche Verbesserung, wie auch alle anderen lokalen R- Faktoren.

In einem weiteren Test (vg. Abbildung 30), der analog der Beschreibung im vorangegangenen Punkt durchgeführt wurde, wurden wiederum alle vorhandenen H $\beta$ -Atompaaubenennungen vertauscht, eine R- Wert- Berechnung durchgeführt, anschließend STEREO ausgeführt und daraufhin erneute R- Werte berechnet.



*Abbildung 30: Unterschiedliche R- Werte bei neuzugeordneten simulierten Daten von HPr. Die gelben Balken zeigen R- Werte bei falsch benannten H $\beta$ - Signalen, die grünen Balken die R- Werte nach der Richtigstellung der Zuordnung durch STEREO.*

Auch hier sieht man eine deutliche Verbesserung der R- Werte bei richtiger stereospezifischer Zuordnung. STEREO war hier in der Lage 40 von 40 vertauschten H $\beta$  zu erkennen und zurückzutauschen. Bei insgesamt vorhandenen 71 diastereotopen Atompaaren ist hier der Effekt auf den globalen R- Wert ebenfalls ausgeprägt, außerdem waren nach dem Rücktausch durch STEREO nahezu keine unzugeordneten Signale vorhanden, was zu einem besseren globalen R- Wert führt.

#### **R- Werte für experimentelle, neu zugeordnete Daten**

Analoge Tests haben wir auch für experimentelle Daten durchgeführt, die automatisch zugeordnet wurden. Abbildung 31 zeigt hier R- Werte bei ursprünglicher Zuordnung und nachdem diese durch STEREO modifiziert wurde. Abbildung 32 die R- Werte bei vertauschen und Richtigstellen aller vorhandenen H $\beta$ .

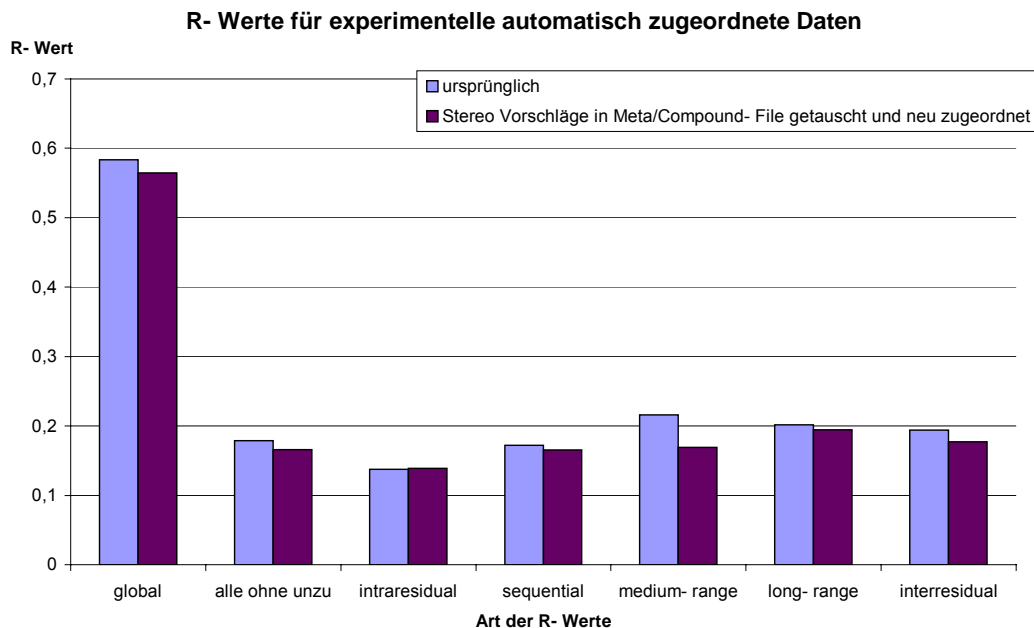


Abbildung 31: Unterschiedliche R- Werte bei automatisch zugeordneten experimentellen Daten von HPr. Die blauen Balken zeigen R- Werte bei ursprünglicher Zuordnung, die violetten Balken die R- Werte nach der Richtigstellung der Zuordnung durch STEREO und einer danach erfolgten automatischen Neuzuordnung.

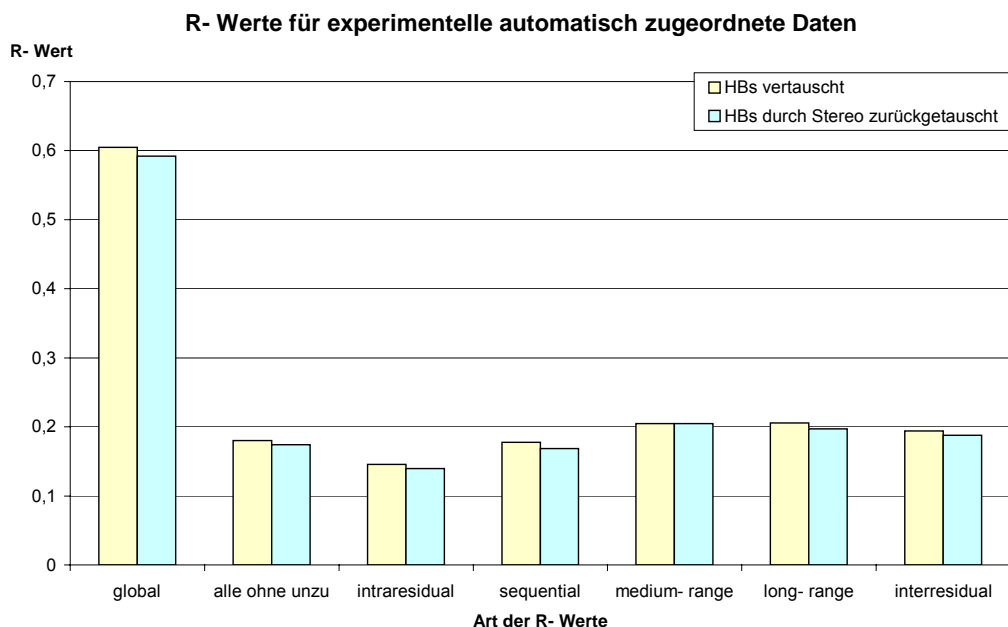


Abbildung 32: Unterschiedliche R- Werte bei automatisch zugeordneten experimentellen Daten von HPr. Die gelben Balken zeigen R- Werte bei falsch benannten H $\beta$ - Signalen, die grünen Balken die R- Werte nach der Richtigstellung der Zuordnung durch STEREO.



Auch hier ist eine signifikante Verbesserung aller R- Werte zu beobachten. Zusammenfassend zeigen die Tests, dass eine richtige stereospezifische Zuordnung zu einer besser definierten räumlichen Struktur, und damit auch zu besseren R- Werten führt. Allerdings sind die beobachteten Effekte relativ klein, insbesondere die Auswirkung auf den, maßgeblich durch unzugeordnete Signale bestimmten, globalen R- Wert. Daher ist es bei dieser Anwendung sinnvoll, den neu definierten R- Wert aus allen vorhandenen, zugeordneten Signalen zu betrachten, der in allen untersuchten Fällen eine Verbesserung bei richtiger stereospezifischer Zuordnung aufweist.

### 3.3.2.3. Anwendungsbeispiele

STEREO ist ein nützliches Programm für die Verbesserung von Proteinstrukturen.

Prinzipiell ist das Programm jederzeit im Strukturfindungsprozess anwendbar. Da die Entscheidungskriterien mittels R- Wert aber auf dem Vergleich von simulierten und experimentellen Daten beruhen, ist es sinnvoll, wenn bei Anwendung des Programms die sequenzielle Zuordnung so weit als möglich abgeschlossen ist.

STEREO arbeitet nur dann sinn- und wirkungsvoll, wenn alle potentiell stereospezifisch zuordenbaren Signale auch als einzelne Signale gekennzeichnet werden, d.h., wenn man nicht z.B. das  $H^{\beta 2}$  und  $H^{\beta 3}$  einer Aminosäure in der experimentellen Liste als zweimal nur  $H^{\beta}$  bezeichnet.

Durch die zusätzliche richtige stereospezifische Zuordnung lassen sich weitere, allerdings durch die relativ geringe Anzahl der Atome, geringe Verbesserungen der Proteinstruktur erzielen, was sich durch einen besseren globalen R- Wert ausdrückt.

Da, wie schon oben erwähnt, die Einflüsse der stereospezifischen Zuordnung auf isolierte NMR- Strukturen eher gering sind, und oft auch die Aussage, die über die zu tauschenden H- Atome getroffen werden kann, sich nur auf diese jeweilige eine Teststruktur bezieht, ist es aussagekräftiger, ein Bündel von NMR- Strukturen zu untersuchen. Dadurch, dass in so einem Fall alle untersuchten Strukturen von ähnlicher Qualität sind, kann man gesicherte Aussagen darüber treffen, welche stereospezifisch benannten Signale richtig zugeordnet sind oder nicht.

Abhängig von der Anzahl der Strukturen in denen ein spezielles Signal als „falsch zugeordnet“ gefunden wurde, lässt sich entscheiden, ob es tatsächlich in der sequenziellen Zuordnung falsch benannt wurde.

Für dieses Anwendungsbeispiel haben wir ein Bündel aus 13 experimentellen NMR-Strukturen des HPr verwendet. Das Bündel wies einen RMSD- Wert von 1,095 Å auf.

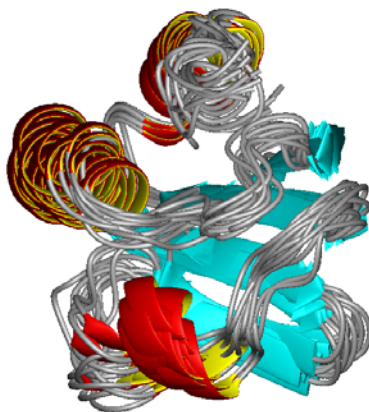


Abbildung 33: Bündel aus 13 HPr Strukturen.

Aus dem Vergleich der in den einzelnen Strukturen gefundenen falschen stereospezifischen Zuordnungen lässt sich bereits mit einer gewissen Wahrscheinlichkeit eine Aussage darüber treffen, welche H- Atom- Bezeichnungen in der eigentlichen Zuordnung getauscht werden sollen, und welche richtig sind. Signale beispielsweise, die in allen 13 Strukturen als falsch identifiziert werden, können nahezu zweifelsfrei als falsch zugeordnet gelten, wogegen Signale, die nur in einer oder zwei Strukturen als falsch betrachtet werden, nicht getauscht werden sollten.

Zunächst wird, wie bei den anderen Teststrukturen auch schon, in jeder einzelnen Struktur des Bündels nach falsch benannten diastereotopen Atomen gesucht.

Bereitet man das Ergebnis weiter auf, so kann man ein Urteil darüber treffen, wie oft wie viele Signale auftreten, was interessant ist, will man unterschiedliche Bündel miteinander vergleichen, oder eine Aussage treffen, bei welchen Signalen man die Benennung tauschen soll. Abbildung 34 stellt das anschaulich dar.

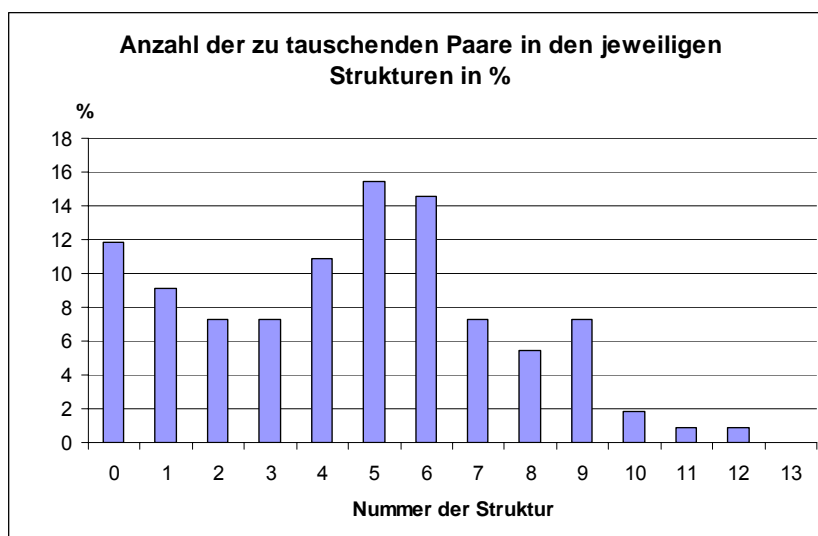
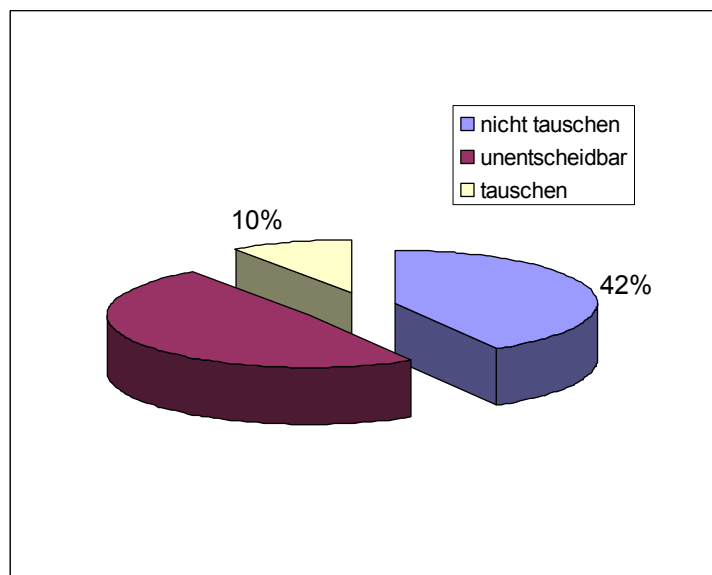


Abbildung 34: Darstellung der Häufigkeit von gleichen zu tauschenden Atomen im Bündel von 13 HPr- Strukturen in %.

Mit den so gewonnenen statistischen Daten kann man entscheiden, wie viele der stereospezifischen Atome in der Zuordnung (Meta-/Compoundfile) getauscht werden sollen. In diesem Fall wurde davon ausgegangen, dass Atome, die in mindestens 70% der Fälle als falsch klassifiziert wurden, auch getauscht werden, und Atome die in bis zu 30% der Fälle richtig waren, nicht getauscht werden. Alle anderen Fälle wurden als unentscheidbar betrachtet.



*Abbildung 35: Statistik über die Aussage, die man über spezielle stereospezifische Atome treffen kann. Berechnet wurde die Anzahl der Atome, die eindeutig getauscht werden sollen aus der Häufigkeit ihres Auftretens im Strukturbündel.*

Das Ergebnis widerspiegelt die Tatsache, dass die meisten  $H\alpha$  ( $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$  etc.) nur statistisch zugeordnet wurden, ohne eine definierte Aussage darüber treffen zu können, welches Signal zu welchem Proton gehört. Dass die Strukturen trotzdem gut definiert sind, liegt daran, dass es sehr viele unterschiedliche Möglichkeiten gibt, eine Modellstruktur zu überprüfen und mit den Messdaten in Einklang zu bringen. Darüber hinaus ist die reine Menge der gefundenen und getauschten Atome, wie schon vorher ausgeführt, im Vergleich zur Gesamtzahl nahezu zu vernachlässigen.

Trotzdem trägt die korrekte stereospezifische Zuordnung zur Verbesserung des Strukturmodells bei. Daher ist es sinnvoll, das hier vorgestellte Programm parallel zum Assignment zu benutzen, um die zeitaufwendige Arbeit der stereospezifischen Zuordnung zu erleichtern. D.h. im Laufe der Arbeit immer wieder die Zuordnung der diastereotopen Atome zu überprüfen.

### 3.3.3. Stereospezifische Zuordnung in Datenbanken- eine Statistik

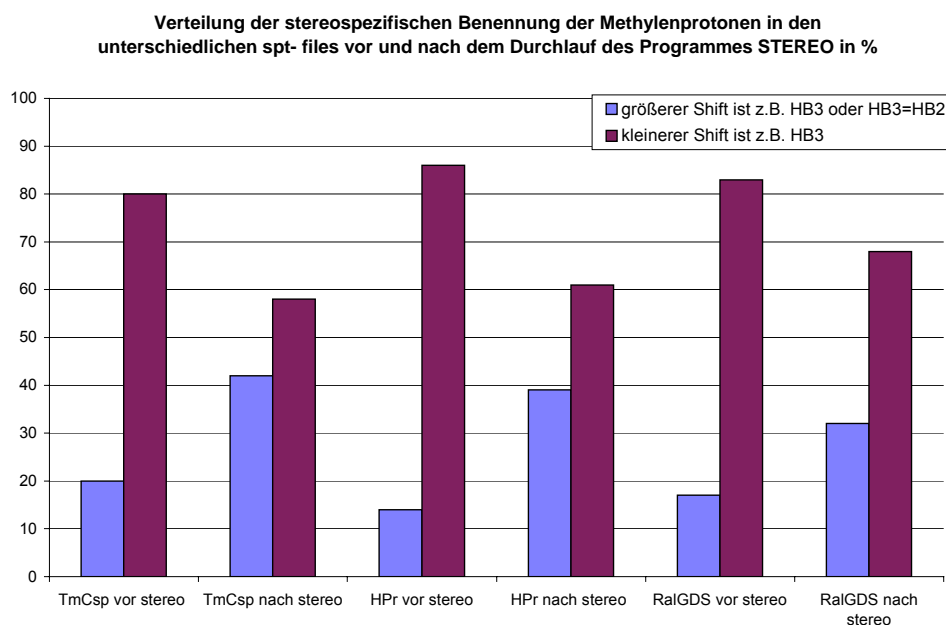
Im Zusammenhang mit den bisherigen Untersuchungen und Ergebnissen taucht die Frage nach möglichen Zusammenhängen von chemischen Verschiebungen und stereospezifischer Zuordnung auf. Gäbe es einen solchen und wäre es möglich diesen zu quantifizieren, würde dies die stereospezifische Zuordnung in Zukunft erheblich erleichtern.

Um diese Zusammenhänge zu untersuchen, haben wir zunächst die originalen stereospezifischen Zuordnungen von TmCSP, HPr und RalGDS ausgewertet.

In Abbildung 36 zeigen die linken Balken Fälle, für die für die Signale mit dem kleineren chemischen Shift die Benennung mit der kleineren Nummer vergeben wurde, die rechten Balken die Fälle, bei denen der größere chemische Shift die Benennung mit der kleineren Nummer trägt. Diese Beziehung wurde vor und nach der Benutzung von STEREO untersucht.

Wie man sieht, besteht ein deutlicher Zusammenhang zwischen kleinerer chemischer Verschiebung und der Namensgebung als „kleinere Nummer“ z.B.  $H^{\beta 3}$ , besonders vor der Überprüfung der Zuordnung mit STEREO. Nach der Benutzung des Programms wird der Zusammenhang weniger signifikant, kann aber in allen Fällen noch beobachtet werden.

Aus dieser Tatsache kann man zweierlei folgern, nämlich zunächst die beobachtete Tendenz von Personen, die Spektren zuordnen, dem kleineren Shift den „kleineren“ Namen zu geben und- wegen der noch vorhandenen Beziehung nach STEREO- dass in mehr als 50% aller Testfälle tatsächlich das Signal mit der kleineren chemischen Verschiebung zu z.B.  $H^{\beta 3}$  gehört.



*Abbildung 36: Stereospezifische Signale in Zuordnungslisten vor und nach der Benutzung von STEREO in %. Vor der Benutzung von STEREO kann man einen deutlichen Zusammenhang zwischen der kleineren chemischen Verschiebung und ihrer Benennung als „niedrigerer Name“ erkennen. Nach STEREO ist diese Beziehung geringer, aber immer noch beobachtbar.*

Diese Feststellung ist sehr interessant und eröffnet die Möglichkeit, auf dieser Basis Wahrscheinlichkeitskriterien zu etablieren, die in bestehende Programme eingebaut werden können um die stereospezifische Zuordnung von Spektren zu erleichtern.

Um diesen beobachteten Zusammenhang noch näher zu untersuchen, haben wir im folgenden 1678  $^1\text{H}$  chemische Shift- Einträge aus der BMRB Datenbank statistisch ausgewertet [Bmr01]. Von diesen waren rund ein Drittel tatsächlich stereospezifisch zugeordnet.

Nachdem offensichtlich einseitig zugeordnete Signale (Signale, die aufgrund des Ambiguity-Codes zwar eindeutig zugeordnet sein sollten, bei denen allerdings in der gesamten Zuordnung alle Signale diesen Ambiguity- Code aufwiesen) aus den Daten eliminiert worden waren, wurde für jede diastereotope Gruppe aus jeder Aminosäure eine Statistik angelegt. Zur besseren Darstellung wurde die Differenz zwischen den beiden chemischen Verschiebungen, die zu einer stereospezifisch zuordenbaren Gruppe gehören, berechnet und diese aufgetragen. Nach einem Fit der Daten mit einer zusammengesetzten Gauß-Funktion stellte sich für die meisten Fälle heraus, dass diese eine leichte Verschiebung zu einer Seite, und hier besonders zur positiven Seite, aufweist. Was bedeutet, dass in den meisten der beobachteten Fälle, tatsächlich der kleinere Shift zu der Benennung als z.B.  $\text{H}^{\beta 3}$  gehört.

Abbildung 37 zeigt die Verteilung der Differenz von zusammengehörenden  $\text{H}^{\beta 2}/\text{H}^{\beta 3}$  Paaren von Asparagin mit einem Gauß Fit, auf 1 normalisiert.

Es ist auch hier offensichtlich, dass das Maximum der Kurve auf der rechten Seite liegt, was bedeutet, dass der Name  $\text{H}^{\beta 3}$  in mehr als 50% der Fälle dem Signal mit dem kleineren Shift zugeordnet wurde<sup>4</sup>.

Abbildung 38, Abbildung 39 und Abbildung 40 zeigen für die Aminosäuren Glutamin, Glutamat und Leucin das gleiche Ergebnis.

Abbildung 41 zeigt für die  $\text{H}^{\beta}$ - Atome von Prolin das umgekehrte Ergebnis, was darauf hindeutet, dass die Beobachtung zumindest keinen systematischen Fehler aufweist.

---

<sup>4</sup> Es ist zu beachten, dass in IUPAC- Nomenklatur der Name  $\text{H}^{\beta 3}$  äquivalent zu der Benennung  $\text{H}^{\beta 3}$  bei einem CNS- oder DYANA output ist, während der Name  $\text{H}^{\beta 2}$  immer  $\text{H}^{\beta 2}$  bleibt.

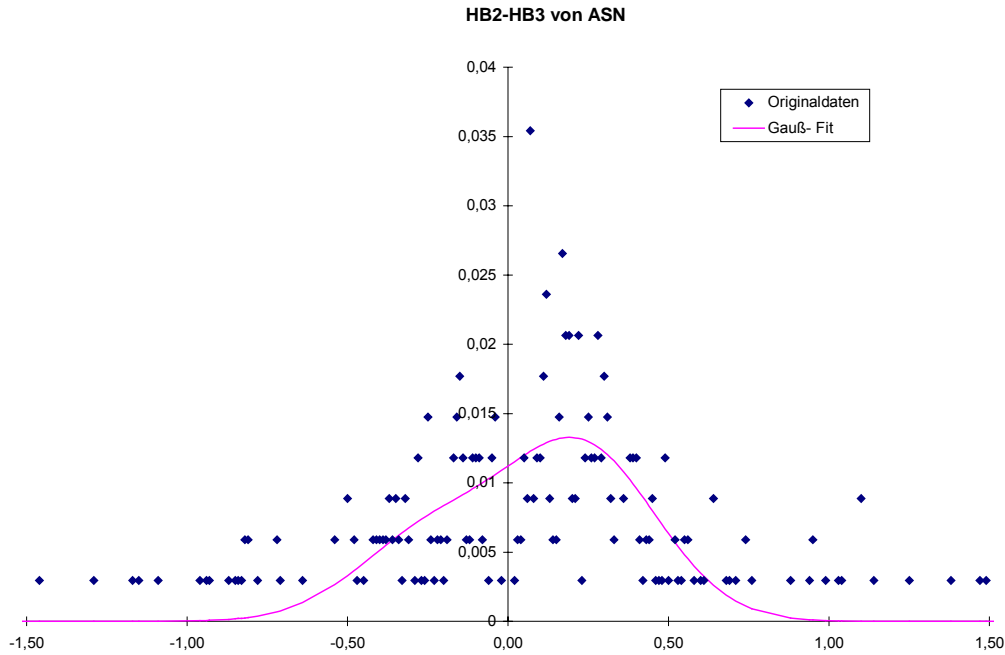


Abbildung 37: Verteilung der Differenz zwischen  $H^{\beta 2}$  und  $H^{\beta 3}$  Shifts von Asparagin. Die absolute Anzahl der Signale wurde auf 1 normiert. Man beobachtet eine leichte Verschiebung des Maximums zur positiven Seite, was bedeutet, dass  $H^{\beta 3}$  in mehr Fällen der kleinere Shift ist.

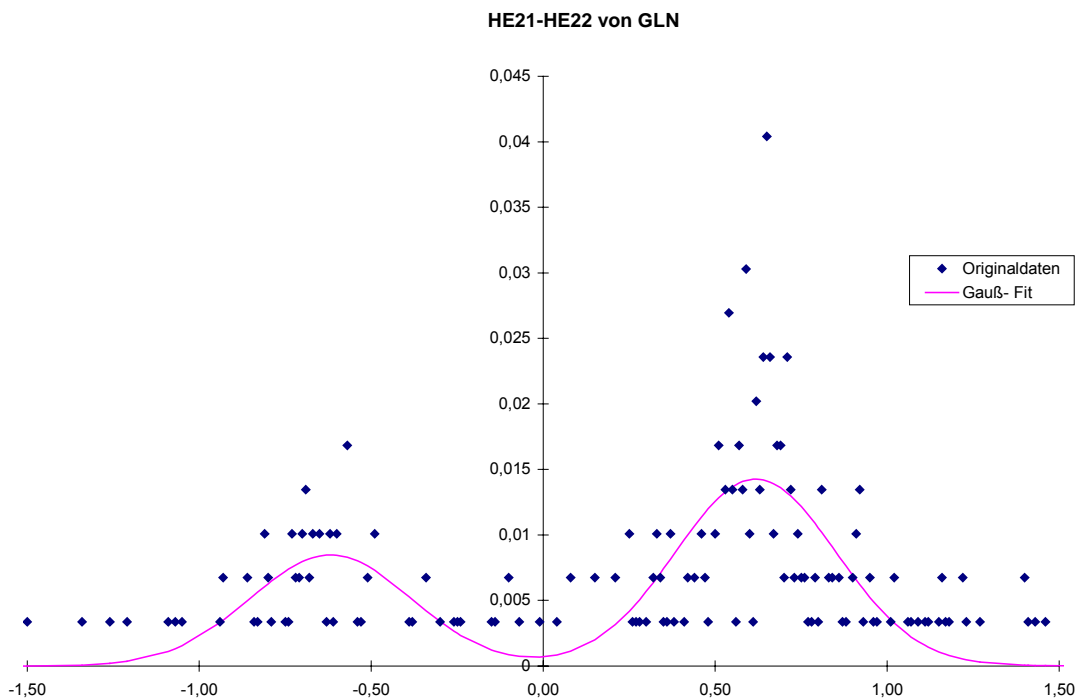


Abbildung 38: Verteilung der Differenz zwischen HE21 und HE22 Shifts von Glutamin. Analog Abbildung 37.

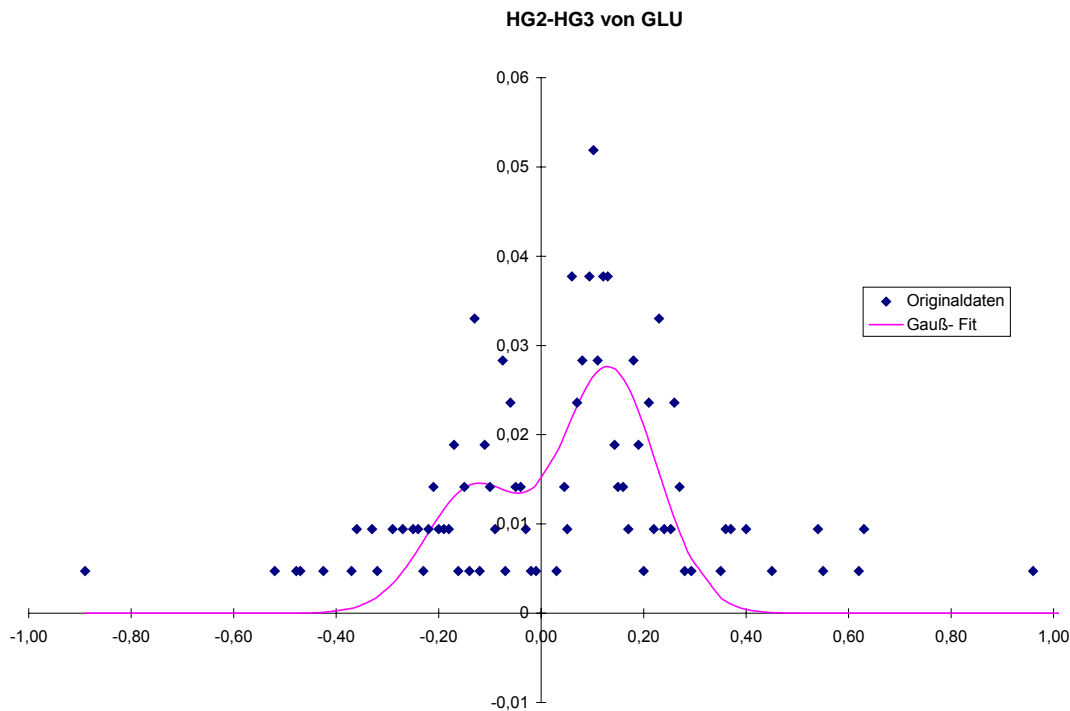


Abbildung 39: Verteilung der Differenz zwischen HG2 und HG3 Shifts von Glutamat. Analog Abbildung 37.

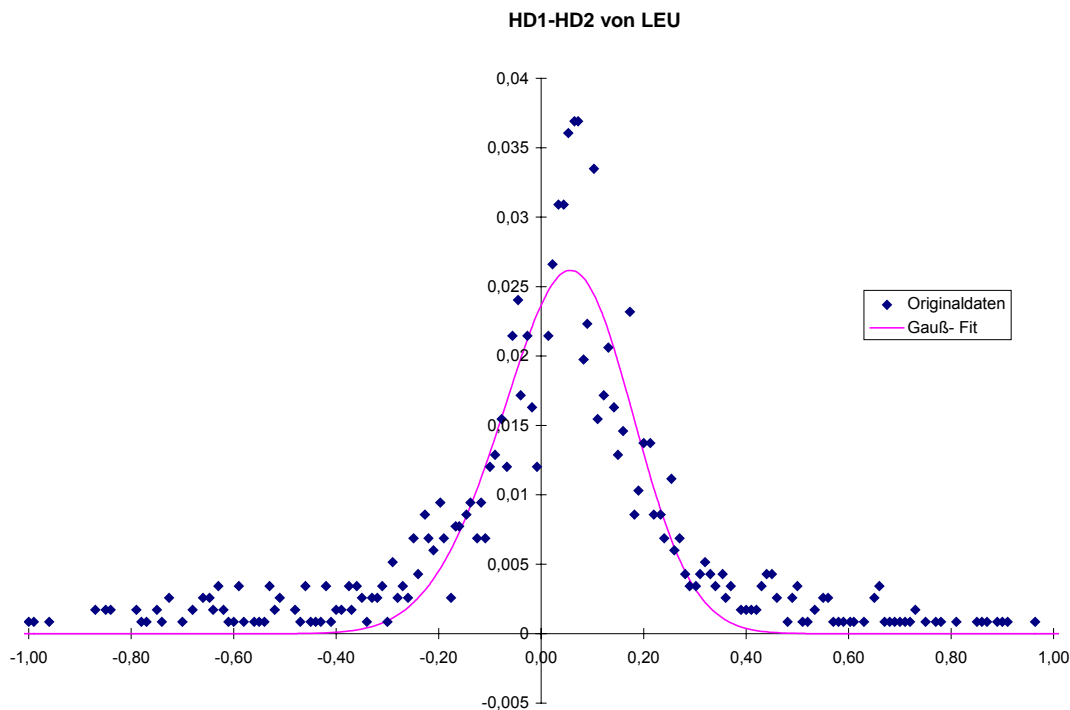


Abbildung 40: Verteilung der Differenz zwischen HD1 und HD2 Shifts von Leucin. Analog Abbildung 37.

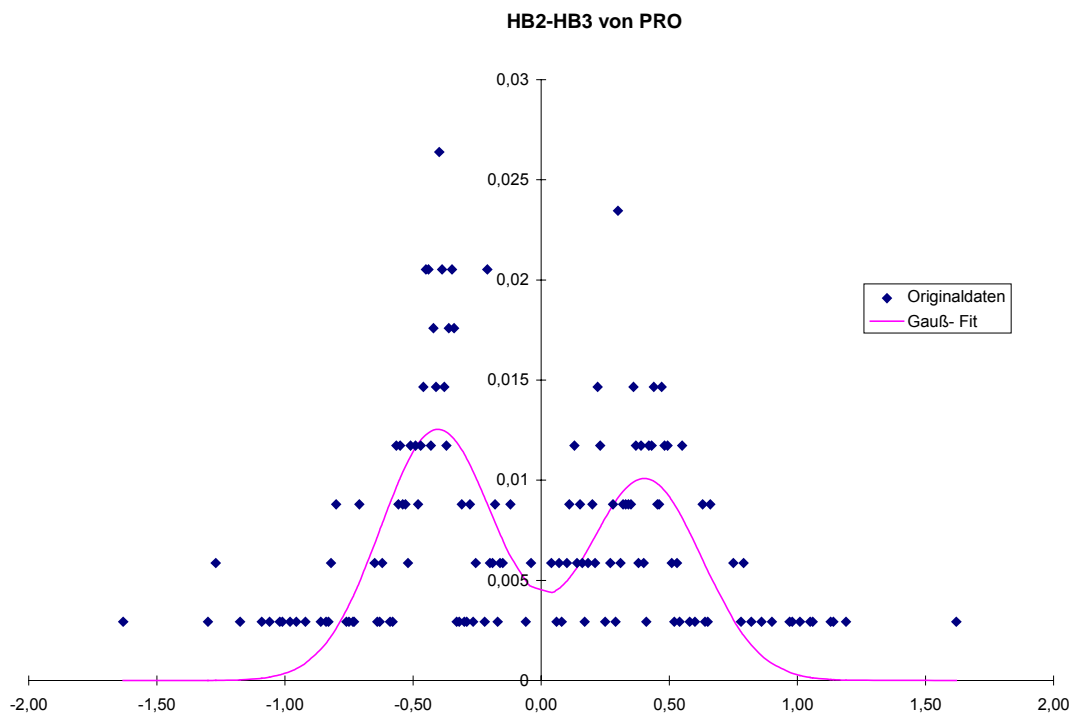


Abbildung 41: Verteilung der Differenz zwischen  $H^{\beta 2}$  und  $H^{\beta 3}$  Shifts von Prolin. Hier ist zu sehen, dass in mehr Fällen, anders als in obigen Abbildungen, der kleinere Shift  $H^{\beta 2}$  ist.

Diese Ergebnisse legen nahe, dass es einen Zusammenhang zwischen der chemischen Verschiebung und der stereospezifischen Zuordnung gibt. Gleichwohl ist es möglich, dass die leichten Verhältnisunterschiede, die beobachtet werden können, aus Fehlern herrühren, die z.B. bei der Konvertierung von DYANA-Format in IUPAC, oder bei der Setzung der Parameter in der Datenbank für „stereospezifisch zugeordnet“ (siehe Abbildung 5: Ambiguity-Code-Definitionen der BioMagResBank.) aufgetreten sind.

Um diese Vermutungen völlig zu erhärten, müssten weitere, gesicherte Daten ausgewertet werden, bei denen die Bestimmung der stereospezifischen Zuordnung mit herkömmlichen Methoden durchgeführt wurden.

## 3.4. Zusammenfassung und Ausblick

### 3.4.1. Zusammenfassung

Wie bereits an anderer Stelle [Gro00, Kir01] beschrieben, wurde in früheren Arbeiten ein Computerprogramm (RFAC) entwickelt, welches die automatische Beurteilung von dreidimensionalen NMR-Proteinstrukturen erlaubt, und durch die Berechnung eines R-Faktors ein zuverlässiges Kriterium für die Qualität dieser Strukturen bietet. Aufbauend auf



dieser Entwicklung haben wir das Programm STEREO dazu erarbeitet, die stereospezifische Zuordnung von diastereotopen Protonen in gegebenen NMR-Strukturen und deren sequenziellen Zuordnungen zu überprüfen. Manuelle stereospezifische Zuordnung ist ein sehr zeitraubender und fehleranfälliger Prozess, der z.B. die experimentelle Bestimmung der Seitenkettenwinkel aus den J-Kopplungskonstanten und NOE-Daten erfordert. Deshalb ist es immer noch üblich, die stereospezifischen Signale statistisch oder gar nicht zu unterscheiden, was einen Verlust an Information bedeutet.

Daher sind besonders in diesem Bereich automatisierte Prozesse, wie STEREO wichtige und nützliche Hilfsmittel, die richtige stereospezifische Zuordnung zu erhalten.

STEREO basiert auf dem Vergleich von experimentellen und simulierten  $^1\text{H}$  NOESY NMR-Spektren bezüglich des stereospezifischen Assignments von diastereotopen Protonen dahingehend, dass für beide Möglichkeiten der stereospezifischen Zuordnung ein lokaler R-Faktor berechnet wird, und nach einem Vergleich, die Möglichkeit mit dem niedrigeren R-Wert gewählt wird. Dieser Mechanismus ist sehr effektiv, da der R-Faktor ein sehr sensibles Werkzeug ist, um strukturelle Änderungen zu erkennen.

Wir haben STEREO erfolgreich an drei mittelgroßen Proteinen getestet: dem Kälteschockprotein (TmCSP) aus *Thermotoga maritima*, dem Histidine containing Phosphorcarrier Protein (HPr) aus *Staphylococcus carnosus*, und der RAS-Bindenden Domäne des RAL Guanosin Nucleotide dissociation stimulators (RalGDS) aus *Homo sapiens*.

Durch Tests mit manuell vertauschten Benennungen stereospezifischer Atome, die generell allesamt erkannt und zurückgetauscht wurden, wurde die einwandfreie Funktion des Programms sichergestellt.

Die Ergebnisse zeigen, dass eine korrekte stereospezifische Zuordnung den R-Wert verringert. Durch die Untersuchung eines Bündels von NMR-Strukturen können statistisch gesicherte Aussagen darüber getroffen werden, ob man in der Zuordnungsliste die Namen eines Paares stereospezifischer Signale tauschen soll, oder nicht. Abhängig von der Anzahl der Strukturen in welchen dieses Signal als falsch zugeordnet klassifiziert wurde, ist es möglich, darüber zu entscheiden.

Zusammenfassend ist STEREO ein einfach zu bedienendes Programm, geschrieben in ANSI-C, das in der Lage ist, falsch zugeordnete stereospezifische Signale mittels experimentellem und rückgerechnetem Spektrum zu identifizieren.

### 3.4.2. Schematischer Programmablauf

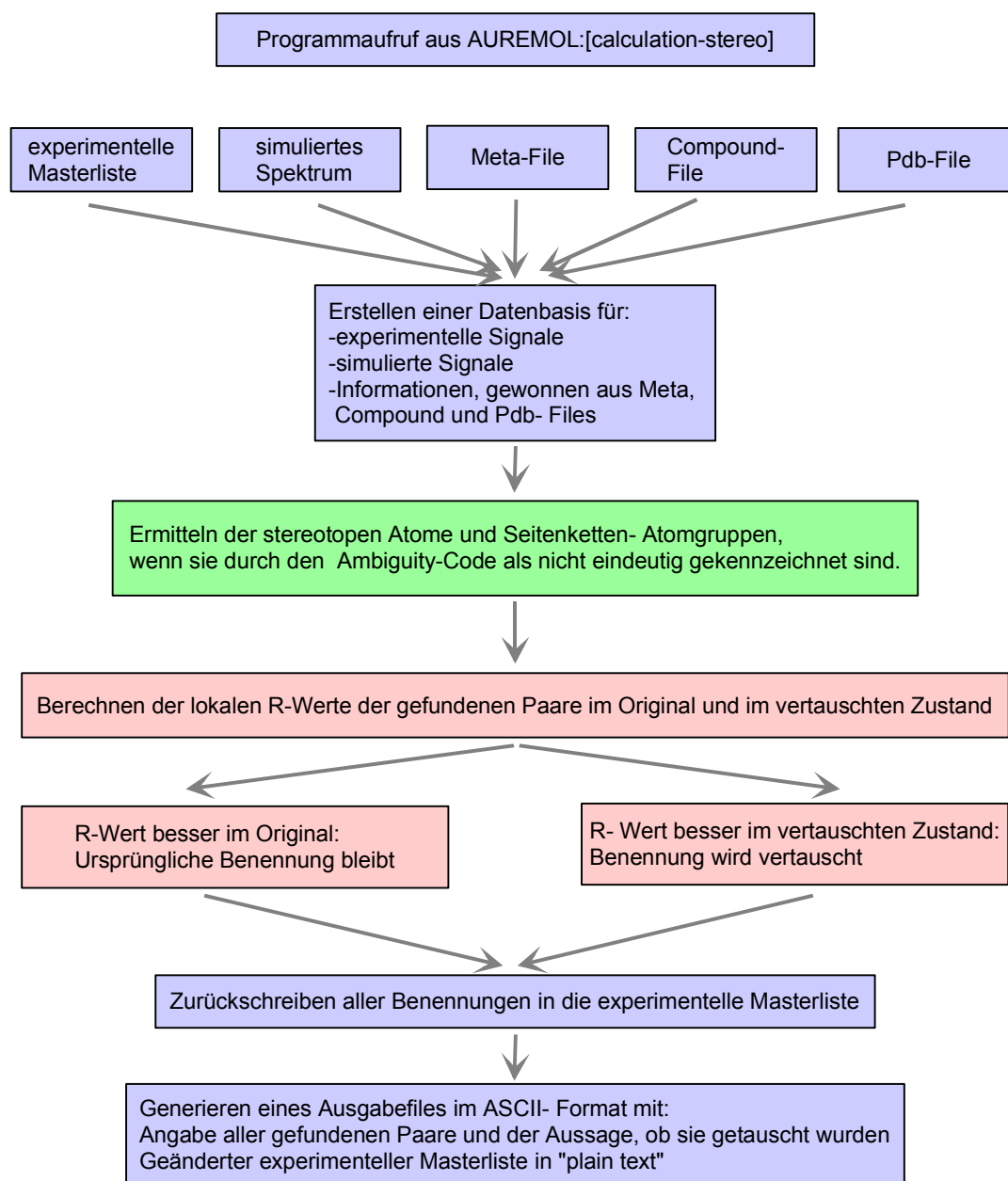


Abbildung 42: Schematischer Programmablauf des Programms STEREO.

### 3.4.3. Ausblick

Für die weitere Zukunft von STEREO ist geplant, zusätzlich zu der Berechnung von R-Werten, Wahrscheinlichkeitskriterien einzuführen.

Diese Wahrscheinlichkeitskriterien ergeben sich aus der Auswertung von ca. 2000 in der PDB- Datenbank vorhandenen  $^1\text{H}$  chemische Shift- Einträge und ihrer Überprüfung auf die unterschiedlichen stereospezifisch zugeordneten Signale.

So lässt sich beispielsweise bei genügender Datenmenge eine statistische Aussage darüber treffen, welches Signal in welcher Aminosäure welche chemische Verschiebung einnimmt, genauer gesagt, ob in einer gegebenen Aminosäure z.B. das  $H^{\beta 2}$  einen niedrigeren oder höheren Wert der chemischen Verschiebung aufweist, als das  $H^{\beta 3}$ .

Dazu wurden bereits Vorarbeiten dahingehend geleistet, dass die Daten erfasst und ausgewertet wurden.

Außerdem wurde für jede Aminosäure eine zusammengesetzte Gauß- Wahrscheinlichkeitsverteilung ihrer stereospezifischen Gruppen erstellt.

Zur weiteren Verifizierung der bereits gewonnenen Erkenntnisse, sollen weitere Daten ausgewertet werden, bei denen die stereospezifische Zuordnung mit herkömmlichen Methoden durchgeführt wurde.

Diese zusätzliche Information wird danach auf der Basis eines Bayes- Wahrscheinlichkeitskriteriums als weiteres Entscheidungsmerkmal in STEREO integriert werden.

## 4. Automatische Zuordnung

### 4.1. Theoretische Betrachtungen, Probleme

#### 4.1.1. Sinn und Zweck

Das NOESY- Spektrum ist momentan noch die wichtigste Informationsquelle für die Bestimmung der dreidimensionalen Proteinstruktur. Um möglichst viele dieser Informationen zu erhalten ist es notwendig 2D oder 3D NOESY Spektren auszuwerten und die dort auftretenden Peaks (NOEs) so vollständig, und vor allem so richtig wie möglich, zu benennen.

Für jedes Paar von Protonen, die räumlich benachbart sind, wird im homonuklearen  $^1\text{H}$  2D NOESY ein Kreuzsignal (im folgenden auch NOE oder Peak genannt) erwartet. Die Intensität der NOEs und damit die Stärke der Wechselwirkung ist in erster Näherung proportional zu  $r^{-6}$ , wobei  $r$  der Abstand der Protonen ist. Die Detektionsgrenze liegt in der Regel bei 0,5 bis 0,6 nm. Besonders wichtig für die Strukturbestimmung sind NOE- Signale von Protonen, die in der Aminosäuresequenz weit auseinander liegen, in der tertiären Struktur aber benachbart sind.

Die Benennung der Spektren geschieht unter Verwendung der sequentiellen Zuordnung des Proteins.

Aus den NOE Daten des NOESY Spektrums ist es möglich Abstandsinformationen zu gewinnen. Die Koordinaten des Peaks (NOEs) im Spektrum entsprechen dabei den Werten der chemischen Verschiebungen der beteiligten Atome. Somit kann man anhand der Zuordnung die an einem NOE beteiligten Atome identifizieren und ihren Abstand ermitteln.

Geht man davon aus, dass dies für sehr viele Signale gelingt (ein 2D NOESY Spektrum enthält in der Regel mehrere tausend NOEs), so ist es auf diese Weise möglich, die Proteinstruktur zu definieren.

Andererseits zeigt sich bereits hier, welcher Arbeitsaufwand zu leisten ist, um die vielen tausend Peaks in der Zuordnung zu suchen, zu identifizieren und schließlich für die Strukturrechnung verfügbar zu machen. Aufgrund verschiedener Probleme, die anschließend erörtert werden, ist dies ein Prozess, der zumindest teilweise, immer noch von Hand durchgeführt werden muss.

Allerdings wäre es, vor allem in Bezug auf die immer weiter anwachsende Menge von Proteinen für die noch keine strukturelle Information verfügbar ist, sinnvoll und notwendig diesen limitierenden Schritt in der Proteinstrukturbestimmung weiter zu automatisieren. Das von uns entwickelte Programm AUNOAS (AUTomated NOe ASSignment) stellt einen Schritt in diese Richtung dar.

### 4.1.2. Probleme

Im Bereich der NMR werden Atome einer Proteinsequenz anhand ihrer chemischen Verschiebung innerhalb der Spektren identifiziert. Am Anfang der Strukturaufklärung werden die zu den einzelnen Atomen gehörenden chemischen Verschiebungen (Shifts) ermittelt. Dieser Prozess der sequentiellen Zuordnung basiert auf verschiedenen homonuklearen und heteronuklearen NMR- Experimenten (3D NOESY, HNCA, CBCAC(O)NH, HCCH TOCSY u.a.).

Das Problem, das sich dabei ergibt ist, dass die Shifts der verschiedenen Atome aus der sequentiellen Zuordnung nicht exakt zu den in den NOESY Spektren gefundenen chemischen Verschiebungen passen, und dass für viele Atome degenerierte chemische Verschiebungen erhalten werden, da der Shift eines Atoms auch von Parametern wie Temperatur, Druck, Salzgehalt der Probe und diversen anderen Parametern abhängt. Dies führt vor allem in Bereichen, in denen die Signale sehr dicht liegen, zu mehreren Möglichkeiten der Zuordnung. Die damit verbundenen Mehrdeutigkeiten können sich in die Strukturbestimmung fortpflanzen, und damit zu schlecht definierten Strukturen führen.

Das ist der Grund dafür, dass Zuordnungen nur bedingt auf Spektren übertragen werden können, die unter leicht abweichenden Bedingungen aufgenommen wurden. Daher ist es aus den eben genannten Gründen momentan schwierig, zu verfolgen, wie sich z.B. das strukturelle Verhalten eines Proteins ändert, wenn äußere Parameter variiert werden.

Allerdings sind es wohl genau diese Dinge, die irgendwann zu einem tieferen Verständnis der Funktion von Proteinen und damit der Funktion von Leben beitragen werden.

## 4.2. Das Programm AUNOAS

### 4.2.1. Funktionsweise

Wertet man von Hand ein NOESY Spektrum aus, so beginnt man in der Regel mit isoliert liegenden NOEs und versucht diese eindeutig zu benennen. Wenn es für diese Signale eindeutige Zuordnungsmöglichkeiten gibt, versucht man damit im weiteren Verlauf überlappende Peaks zu benennen. Dieser Prozess ist sehr fehleranfällig, da es Menschen naturgemäß sehr schwer fällt viele sich gegenseitig beeinflussende Faktoren zu berücksichtigen.

Allerdings ist es mit geeigneten Algorithmen möglich, diesen sich selbst beeinflussenden Prozess computergestützt ablaufen zu lassen.

Der Algorithmus, der dem Programm AUNOAS zugrunde liegt, stellt im Wesentlichen die menschliche Vorgehensweise nach und versucht, von sicheren Erkenntnissen auf noch „unsichere“ Erkenntnisse zu schließen. Dieser Prozess wird iterativ wieder auf sich selbst rückgekoppelt und optimiert so ständig seine Eingabeparameter. Am Ende wird eine auf der

ursprünglichen Zuordnung basierende neue Zuordnung erzeugt, die das zu benennende Spektrum (auch unter Einbeziehung von Strukturdaten) so optimal wie möglich erklärt. Dieser iterative, rückgekoppelte Optimierungsprozess sorgt dafür, dass das Programm gut mit Daten arbeiten kann die entweder verrauscht sind, oder von den Erwartungswerten der anfänglichen Zuordnung abweichen. Bei der Programmarchitektur haben wir großen Wert auf Flexibilität in der Anwendung gelegt. D.h. es sollte durch geeignete Wahl der Funktionsparameter möglich sein, sowohl Spektren auszuwerten für die noch keine, oder nur eine unsichere Strukturvorstellung existiert, als auch das Verhalten von Proteinen, deren Struktur bekannt ist, unter z.B. Veränderungen der äußeren Bedingungen zu analysieren.

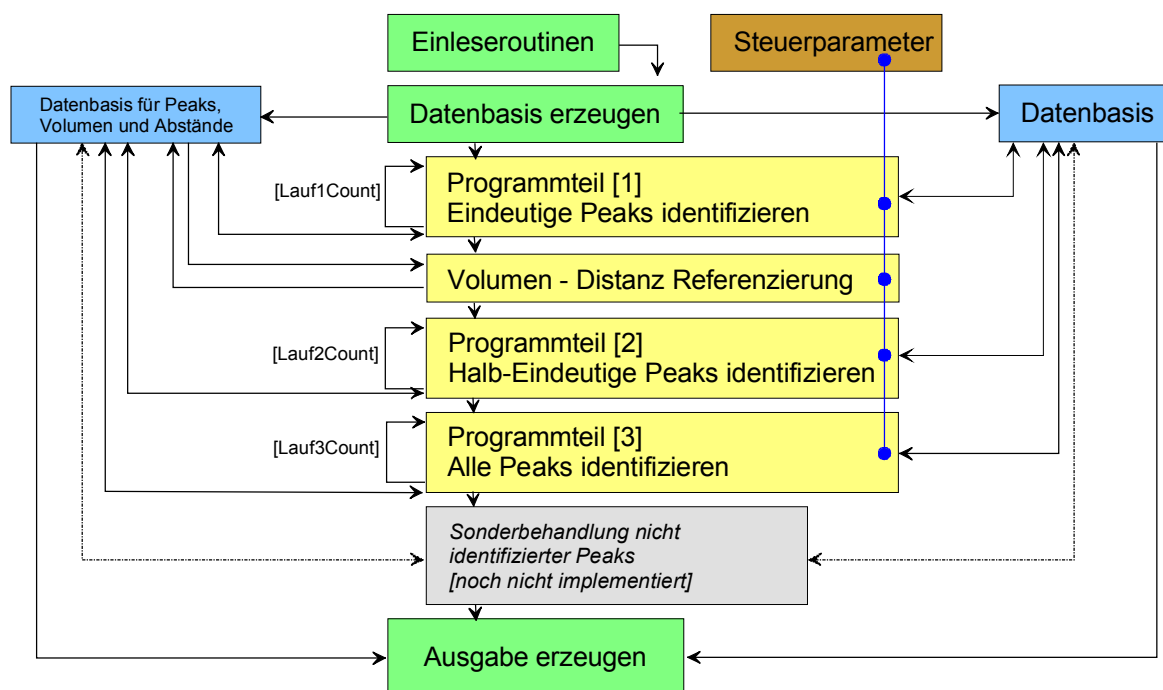


Abbildung 43: Stark vereinfachtes Diagramm, das den Datenfluss während der Spektrenbenennung durch AUNOAS darstellt. Deutlich zu sehen ist, wie alle Programmteile auf die gleiche Datenbasis zugreifen und diese auch ändern.

Abbildung 43 stellt eine stark vereinfachte Übersicht des Programmablaufes dar. In einem ersten Schritt werden nach dem Einlesen und Aufbereiten der Daten isoliert liegende Signale mit nur einer eindeutigen Zuordnungsmöglichkeit benannt, im zweiten Schritt Signale, die in mindestens einer Koordinate eindeutig sind, und schließlich in einem letzten Schritt werden alle noch übrigen Signale mittels eines Qualitätsbeurteilungskriteriums überprüft und gegebenenfalls benannt. Alle Programmteile greifen auf die gleiche Datenbasis zu und verändern diese auch. Durch diesen iterativen Änderungsprozess passen sich die Einträge in der Datenbasis immer besser dem experimentellen Spektrum an. Steuerparameter, die für jeden Programmteil verfügbar sind, sorgen für den reibungslosen Ablauf. Im Folgenden werden die Datenbasis, alle Programmteile, sowie die Programmparameter eingehend beschrieben.

## 4.2.2. Datenbasis

### 4.2.2.1. Verwendete Daten

#### ***Spt- File oder meta- File und compound- File:***

Enthält die Zuordnung des Proteins. Die Zuordnung sollte so vollständig wie möglich sein. Die Daten können sowohl im CNS Format als auch im IUPAC Standard eingelesen werden.

Wichtig ist jedoch, dass sowohl im pdb-, als auch im spt-/meta- und im RELAX- output- File dieselbe Benennungskonvention verwendet wird.

Bei Verwendung des Programms aus AUREMOL heraus ist zu beachten, dass alle Daten nur im IUPAC- Format verarbeitet werden, und kein spt- File Format mehr verwendet werden kann.

#### ***Pdb- File:***

Um ein experimentelles NOESY Spektrum zuzuordnen, ist es vorteilhaft, strukturelle Information mitzuverwenden. Diese Information ist im pdb- File anhand der Koordinaten der einzelnen Atome des Proteins gespeichert.

#### ***Masterliste des zuzuordnenden Spektrums:***

Die Information für das zuzuordnende Spektrum stammt aus der Masterliste. Aus dieser Liste werden die chemischen Verschiebungen der Peaks, ihre Volumen, die von AUREMOL vergebene Signalklassifizierung sowie die Information, ob es sich um ein 2D oder ein 3D Spektrum handelt, gewonnen.

#### ***Optional: rückgerechnetes RELAX- output- File:***

Mit der "backcalculation"- Funktion in AUREMOL ist es möglich, ein Spektrum anhand Strukturdaten zu simulieren. Das Ausgabe- File dieses Prozesses kann zusätzlich dazu verwendet werden, den Zuordnungsprozess zu verfeinern, da die dort berechneten Volumen unter Berücksichtigung der Spindiffusion erzeugt werden.

Das Format der benutzen Eingabedaten ist dem Kapitel 3.2.2.2 zu entnehmen.

### 4.2.2.2. Interne Datenaufbereitung

Das Programm AUNOAS verwendet die oben aufgeführten Eingabe- Files und organisiert die eingelesenen Informationen in einer gemeinsamen Datenbasis. Dabei werden den einzelnen Atomen sowohl ihre chemischen Verschiebungen, als auch ihre räumlichen Koordinaten in der Struktur zugewiesen. Diese Datenbasis enthält ebenfalls ein Feld für die Limitierung, die bei der Spektrenbenennung durchgeführt wird, und ein weiteres Feld, in dem angegeben wird, wie oft das entsprechende Atom im Verlauf der Zuordnung als eindeutiger

Bestandteil eines Peaks identifiziert wurde. Außerdem wird gespeichert, welcher Aminosäure das Atom zugehört.

Ununterscheidbare Atome wie z.B. HE\* (HE1, HE2, HE3) werden mit ihren Koordinaten als einzelne Atome mit gleichem Namen (HE) aufgeführt.

**Struktur der internen Datenverwaltung:**

Atom	Art	AS	Shift	Limit	x	y	z	Gefunden
HB1	MET	1	1,61	0,03	48,59	-9,84	-3,58	12
HB2	MET	1	1,58	0,03	48,87	-8,33	-2,72	13
HG1	MET	1	2,04	0,03	47,76	-9,04	-0,84	20
HE	MET	1	2,10	0,06	44,46	-11,32	-1,44	10
HE	MET	1	2,10	0,06	45,99	-11,49	-0,58	10
HE	MET	1	2,10	0,06	45,77	-12,25	-2,15	10

Abbildung 44: Darstellung der Organisation der verwendeten Datenbasis.

Durch das Zusammenführen der einzelnen Informationen aus den unterschiedlichsten Quellen und der Organisation in einer gemeinsamen Tabelle wird einerseits die Zugriffszeit verkürzt, andererseits die Übersichtlichkeit enorm vergrößert.

### 4.2.3. Programmbeschreibung und -ablauf

#### 4.2.3.1. Programmaufruf

Das Programm AUNOAS wurde in ANSI- C geschrieben und ist mittlerweile im Programmpaket von AUREMOL integriert. Dort sind alle nicht vom Benutzer anzugebenden Parameter festgelegt, und können nicht verändert werden. Des weiteren wurde es in Angriff genommen, die Berechnung der internen Qualitätswerte mittels einer Bayes-Wahrscheinlichkeit durchzuführen. Für speziellere Anwendungen ist jedoch zusätzlich noch eine „standalone“- Version verfügbar, bei der je nach Bedarf in den Code eingegriffen und diverse Parameter verändert werden können. Alle im weiteren Verlauf dieser Arbeit gezeigten Ergebnisse wurden mit dieser Version erzeugt.

Diese „standalone“- Version wird aus der Kommandozeile eines Dos- oder UNIX- Fensters aufgerufen , wobei dabei ein Hilfe- Eintag mitausgegeben wird, wie das Programm zu bedienen ist. Diese Version arbeitet nicht mit rückgerechneten Spektren.

In AUREMOL findet sich das Programm AUNOAS unter dem Pull- down Menü "Calculation".



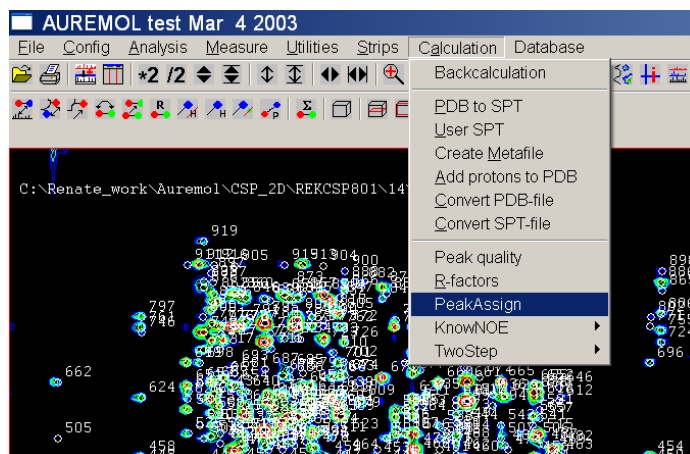


Abbildung 45: Programmaufruf aus AUREMOL heraus unter dem Menüpunkt „calculation“.

Dort wird mit dem Punkt PeakAssign eine Eingabemaske gestartet, in der mit den zu verwendenden Files auch die oben beschriebenen Parameter angegeben werden müssen. Zunächst muss der Benutzer ein 2D oder 3D NOESY- Spektrum in AUREMOL auswählen. Zusätzlich wird ein File mit der sequenziellen Zuordnung (spt- oder meta- File) und ein File mit den strukturellen Informationen (pdb- File) benötigt.

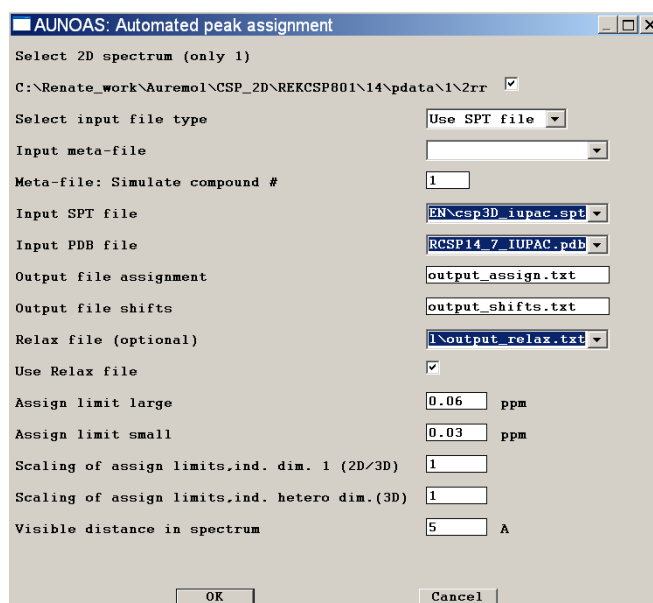


Abbildung 46: Eingabemaske des Programmes AUNOAS. Angegeben werden müssen die entsprechenden Dateien und die in 4.2.4 besprochenen Parameter.

Optional kann ein rückgerechnetes (simuliertes) Spektrum verwendet werden (dessen output- File vorher mittels der "backcalculation"- Funktion in AUREMOL berechnet werden muss), das dann verwendet wird, um die automatische Zuordnung mit Hilfe der

rückgerechneten Volumen, anstelle der aus den experimentellen Daten gewonnenen Distanz, zu erzeugen.

Es ist zu beachten, dass alle verwendeten Dateien derselben Benennungskonvention folgen, am Besten alle im IUPAC- Format sind.

Zusätzlich zur Aktualisierung der Masterliste nach dem Programmlauf, werden zwei textbasierte Ausgabe- Files erzeugt, teilweise mit zusätzlichen Informationen. Diese Ausgabe- Files können in anderen Programmen, beispielsweise EXCEL weiterverwendet werden.

Das **output- File assignment** enthält die Ergebnisse der Zuordnung, die im Programm berechnete Distanz, das Volumen, und den intern berechneten Qualitätswert.

Das **output- File shifts** enthält die an das aktuelle Spektrum angepasste (veränderte) sequenzielle Zuordnung, in einem Format, das dazu benutzt werden kann, mit AUREMOL ein neues **meta- File** zu erzeugen.

Die im Folgenden anzugebenden Parameter werden später (siehe Abschnitt 4.2.4, S.71 ff) besprochen.

Sind die Voreinstellungen getroffen, beginnt der Zuordnungsprozess. Ein Statusbalken informiert den Benutzer über den Fortschritt in der Zuordnung.

#### 4.2.3.2. Einleseroutine

Nach dem Programmstart werden die angegebenen Files eingelesen und die entsprechenden Starteinstellungen getroffen.

#### 4.2.3.3. Datenaufbereitung

Ausgehend von der Zuordnung im spt (meta)- File und den Atomkoordinaten aus dem pdb- File wird die Datenbasis erzeugt. Dabei werden ununterscheidbare Atome wie hier z.B. HE\* als einzelne Atome mit dem jeweiligen Shift behandelt.

HE	1	2,097	0,06	44,46	-11,32	-1,44	5
HE	1	2,097	0,06	45,99	-11,49	-0,58	5
HE	1	2,097	0,06	45,77	-12,25	-2,15	5

*Abbildung 47: Behandlung ununterscheidbarer Atome als einzelne Atome mit der selben chemischen Verschiebung, am Beispiel des HE\*.*

Für jeweils ein Atom wird dann in der internen Datenbasis ein Eintrag vorbereitet, der den Atomnamen, die zugehörige Aminosäure, den zugeordneten Shift, die Limitierung (anfangs BigLimit), die x, y, z- Koordinaten sowie den Parameter Startcount (hier 5) enthält.

Im Programmverlauf wird in dieser Datenbasis die Limitierung geändert, für jede Identifizierung eines Shifts wird dessen Wert neu, anhand einer Mittelwertbildung, errechnet und der Counter für die Anzahl der Identifizierungen wird erhöht.

Nach dem Einlesen und Aufbereiten der Daten beginnt der eigentliche Benennungsvorgang. Dabei werden in mehreren Durchläufen jeweils alle Peaks überprüft und gegebenenfalls benannt. Diese Durchläufe sind ihrerseits in drei Programmteilen zusammengefasst, die unterschiedliche Aufgaben bei der Peakbenennung erfüllen. Die Anzahl der Durchläufe in den verschiedenen Programmteilen wird durch die Parameter Lauf1Count, Lauf2Count und Lauf3Count festgelegt.

Zwischen den Programmblöcken [1] und [2] befindet sich ein „Zwischenprogramm“, das die Volumina des experimentellen Spektrums in Distanzen umrechnet. Die Basis für diese Umrechnung wird im Programmteil [1] erzeugt.

#### 4.2.3.4. Bearbeiten und Benennen der eindeutigen Peaks

##### **Programmteil [1]**

Im ersten Programmteil werden isoliert liegende und eindeutig zu benennende Peaks gefunden und benannt. Für diese Peaks sind damit auch die exakten Shifts der beteiligten Atome im experimentellen Spektrum bekannt. Der ermittelte Shift ( $Shift_{neu}$ ) wird nun in der Datenbasis zu dem dort bereits vorhandenen zugeordneten Shift ( $Shift_{alt}$ ) addiert und dann einer gewichteten (MasterQual: Qualitätswert aus Masterliste) Mittelwertbildung unterzogen. Der so neu ermittelte Shift ersetzt nun in der Datenbasis den ursprünglichen Wert ( $Shift_{alt}$ ) und der Counter (Count), der die Anzahl der Identifizierungen für das jeweilige Atom enthält, wird um  $MasterQual^5$  erhöht ( $Count = Count + MasterQual$ ).

Am Anfang der Berechnungen enthält Count den Wert Startcount aus der Parameterdefinition (vgl. 4.2.4).

$$Shift = \frac{(Shift_{alt} \cdot Count) + (Shift_{neu} \cdot MasterQual)}{Count + MasterQual}$$

*Gleichung 7: Berechnen der neuen Shiftposition*

---

<sup>5</sup> *MasterQual=1 wenn [MasterQualitaet]=nein, bzw. wenn Peak in Masteliste als "echter" Peak gekennzeichnet ist*

**Anmerkung:** Das wiederholte Abarbeiten der Peakliste führt zur Annäherung der Shifts in der Datenbasis an den Wert des experimentellen Spektrums. In den ersten Durchläufen ist der Beitrag eines identifizierten Peaks zur Mittelwertbildung noch relativ groß. Während dieser ersten Durchläufe hängt der momentane Stand der Datenbasis auch noch von der Abfolge der Peaks in der Liste ab.

Um diesen Einfluss abzuschwächen wird die ursprüngliche Zuordnung anfangs höher gewertet (Startcount) und die Limitierung des Fangradius bleibt ebenfalls während der ersten Läufe (Laufnochange) unverändert. Der Beitrag eines Peaks wird mit der Anzahl der Durchläufe immer geringer. Der Wert in der Datenbasis nähert sich also asymptotisch an den Wert im Experiment an.

Würde man die Mittelwertbildung am Ende eines Durchlaufes durchführen so würde während dieses Durchlaufes nur die ursprüngliche Zuordnung verwendet. Damit würde in diesem Durchlauf der Einfluss auf das gesamte Spinsystem erst wieder im nächsten Lauf berücksichtigt. Deshalb wären mehr Durchläufe und damit eine längere Programmlaufzeit erforderlich.

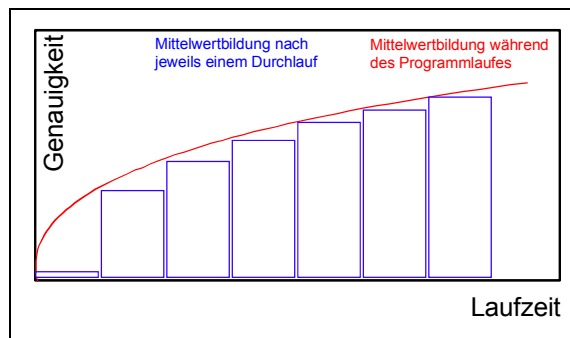


Abbildung 48: Annäherung der Shifts in der Datenbasis an den experimentellen Wert.

**Der Ablauf für den Programmteil [1] stellt sich also wie folgt dar:**

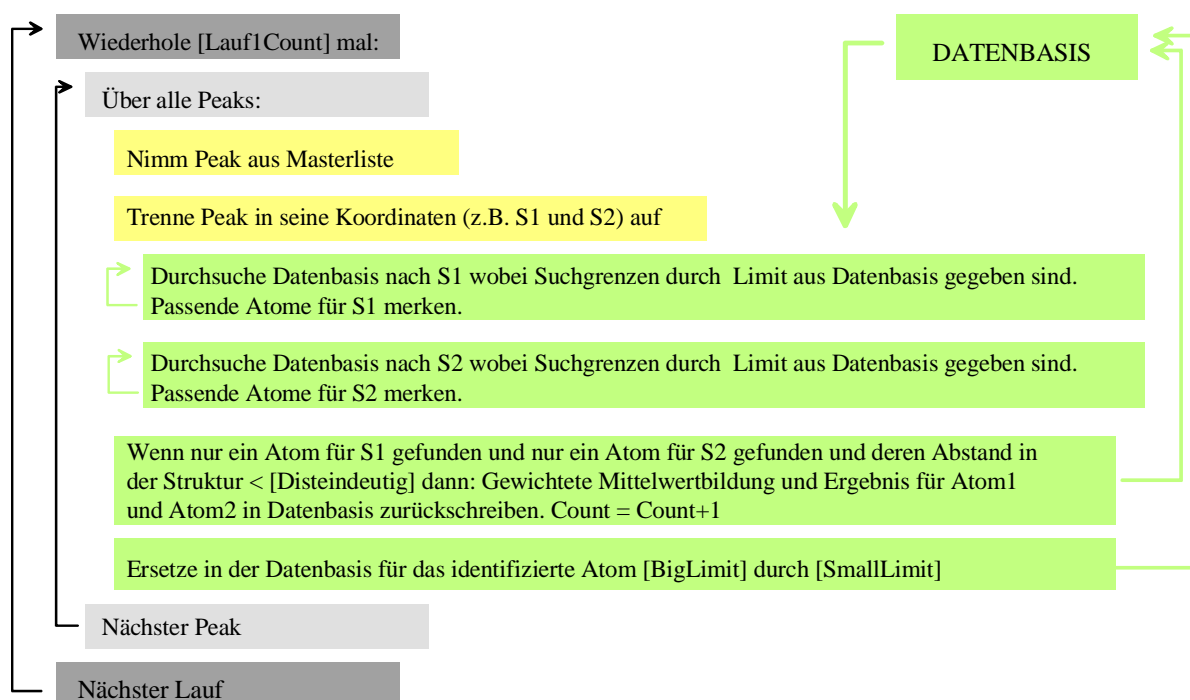


Abbildung 49: Schematische Darstellung des Programmteils [1].

Des weiteren wird jedem eindeutig bestimmten Peak in der Datenbasis der strukturelle Abstand, berechnet aus der Struktur, zugeordnet.

#### 4.2.3.5. Volumen - Distanz Referenzierung

Aufgabe dieses Programmteiles ist es, anhand der im Programmteil [1] für sichere NOEs ermittelten strukturellen Abstände einen Eichparameter zu finden.

W1	AS 1	W2	AS 2	Shift 1	Shift 2	DIST	Volumen	Qualität
1				9,430	9,785		134591	0,99
2				9,324	9,439		75968	0,44
HN	17	HN	25	9,313	9,785	2,55	580411	1,00
HN	6	HN	18	9,242	9,350	4,12	454124	1,00
5				9,078	9,245		121454	1,00
6				8,996	9,204		685691	1,00
7				8,996	9,433		63900	1,00
8				8,820	9,257		526014	1,00
9				8,808	9,198		160397	0,52
10				8,726	9,045		236557	0,96

*Abbildung 50: Darstellung der Datenbasis zum gegenwärtigen Zeitpunkt. Für mehrere Signale ist eine Zuordnung getroffen, diesen werden ihre räumlichen Distanzen zugeordnet, und anschließend werden sie herangezogen, um das Volumen in Distanzen umzurechnen.*

Da die räumliche Distanz zweier bekannter Atome aus der gegebenen Struktur leicht berechnet werden kann, ist es zu diesem Zeitpunkt möglich, zunächst den bekannten Atomen ihre Distanz zuzuordnen, und anschließend, in einem weiteren Schritt durch Mittelwertbildung einen Parameter zu errechnen, der die in diesem Spektrum gegebenen Volumina in Distanzen umrechnet.

Anhand dieses Eichparameters und den Parametern ExpMinDist und ExpMaxDist (maximale und minimale Entfernung zweier Atome, die im Spektrum sichtbar sind) wird jedem Peak ein struktureller Abstand zugewiesen.

Wird ein rückgerechnetes Spektrum für die Zuordnung mitverwendet, so ergibt sich die Berechnung der Distanzen aus den darin enthaltenen Werten (siehe Abschnitt 4.2.3.7).

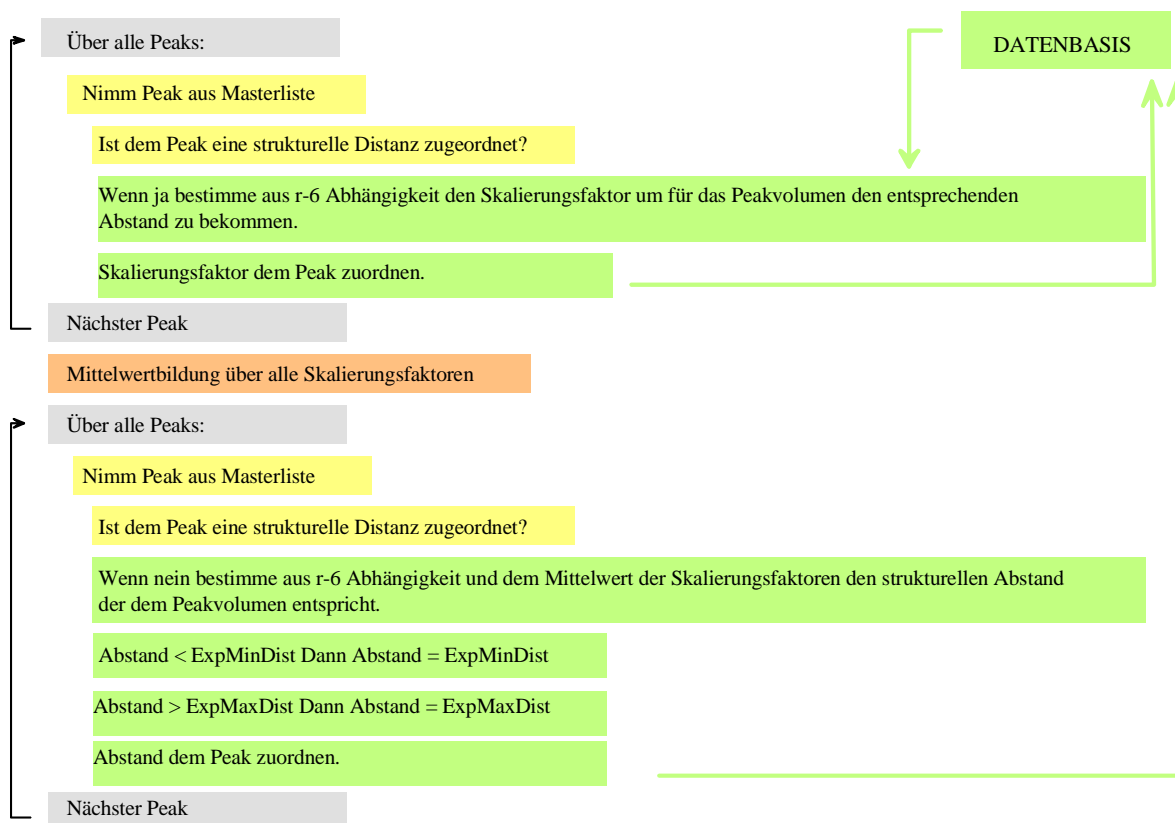
**Programmablauf:**

Abbildung 51: Schematische Darstellung des Ablaufs der Distanzberechnung aus den Volumen des experimentellen Spektrums.

Nach dem Ablauf des „Zwischenprogramms“ sind allen Peaks entsprechende strukturelle Abstände zugeordnet. Natürlich sind diese Abstände aufgrund von z.B. Spindiffusion, oder der Tatsache dass manchmal mehrere Peaks überlagert sind, fehlerbehaftet. Trotzdem stellen diese Abstände ein wertvolles Instrument dar, um in den weiteren Programmteilen zu entscheiden, welcher Peak von welchen Atomen aufgebaut wird.

Tests während der Entwicklungsphase des Programms haben auch gezeigt, dass die errechneten Skalierungsfaktoren erstaunlich gut clustern und die errechneten Werte für die Distanzen der strukturellen Wirklichkeit sehr genau entsprechen.

Betrachtet man eine solche errechnete Distanzverteilung für ein Modellprotein (hier: Saratin) so findet sich eine annähernde Gauß- Kurve wieder, was auch für die reale Verteilung der Distanzen zu erwarten ist.

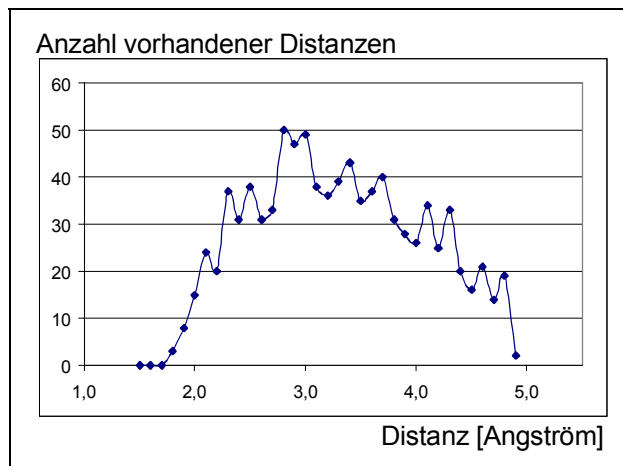


Abbildung 52: Abstandsverteilung für Saratin; errechnet aus einem 2D NOESY Spektrum

Macht man sich Gedanken über die gegebenenfalls auftretenden Fehler, so kann man davon ausgehen, dass sie ca. 20% nicht überschreiten, was für diese Anwendung ein akzeptabler Wert ist.

**Fehlerabschätzung:** 3 Atome bilden gleichberechtigt einen NOE; Programm geht von 1 Atom aus:

Abstand für 1 Atom:	Volumen ≈ 2 Angström	Abstand für 3 Atome:	Volumen/3 ≈ 2,4 Angström	Fehler:	ca. 20 %
---------------------	-------------------------	----------------------	-----------------------------	---------	----------

#### 4.2.3.6. Anpassung von simulierten an experimentelle Volumen

Aufgabe dieses Programmteiles ist es, simulierte Volumen aus der Rückrechnung durch eine Abstandbasierte Skalierung so an die experimentellen Volumen der Spektren anzupassen, dass diese später ohne Vorfaktoren miteinander verglichen werden können.

Der Skalierungsfaktor berechnet sich nach:

$$Scale = \left( \frac{\sum \left( \left( Vol_{exp} \cdot Vol_{sim} \right)^{-\frac{1}{6}} \right)}{\sum \left( \left( Vol_{sim}^2 \right)^{-\frac{1}{6}} \right)} \right)^{-6}$$

Gleichung 8: Berechnung des Skalierungsfaktors

Außerdem sorgt diese Routine dafür, dass verschiedene Spektren in verschiedenen Programmdurchläufen immer gleich skaliert werden.

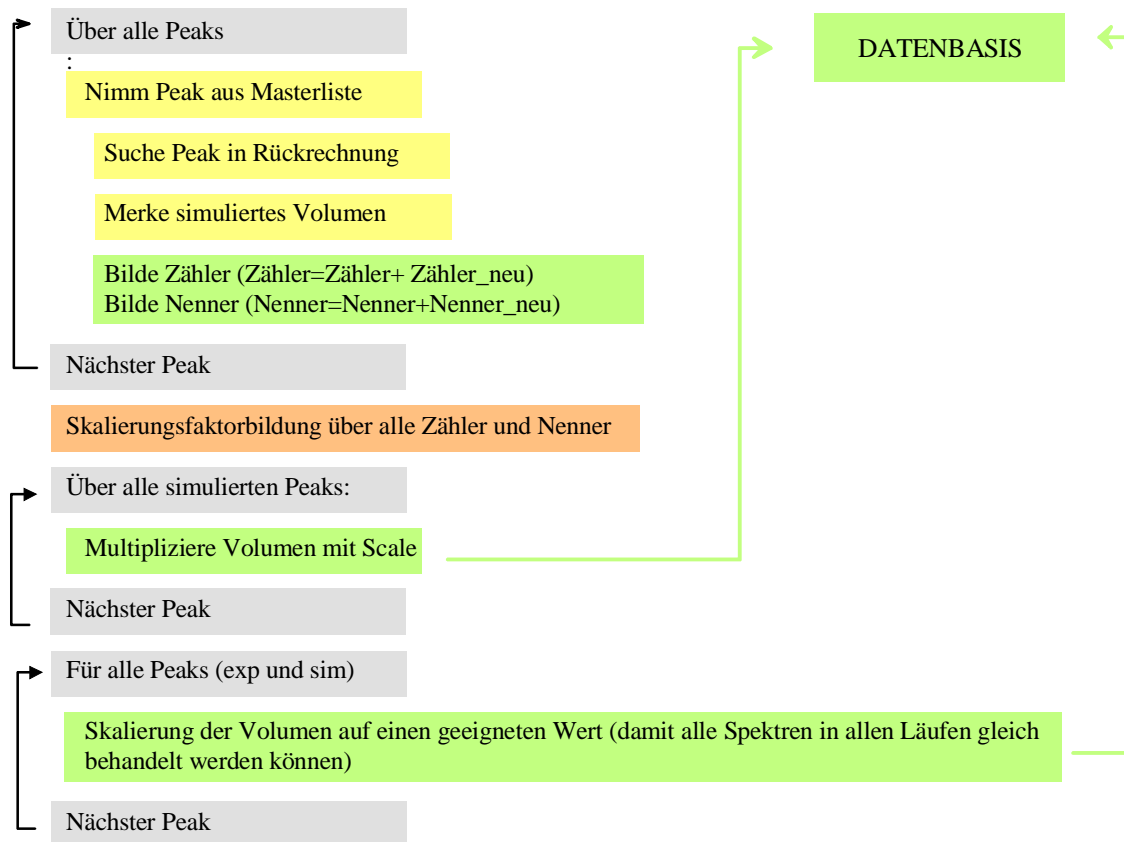
**Programmablauf:**

Abbildung 53: Schematische Darstellung der abstandbasierten Skalierung von simulierten und experimentellen Spektren.

#### 4.2.3.7. Peaks mit einer eindeutigen Koordinate bearbeiten

##### Programmteil [2]

Dieser Programmteil ordnet ausgehend von bereits gewonnenem Wissen anhand der nun schon teilweise an das experimentelle Spektrum angepassten Zuordnung so viele Peaks wie möglich zu, und verbessert die experimentspezifische Zuordnung weiter.

Im zweiten Schritt des Benennungsvorganges werden wiederum alle Peaks in Lauf2Count Durchläufen überprüft.

Im Unterschied zu Programmteil [1] wird nun aber zu jeder möglichen Peak- Erklärung ein Qualitätswert berechnet. Einem experimentellen Peak wird die Bezeichnung zugeordnet die den besten Qualitätswert aufweist, und als weiteres Kriterium muss der Peak zusätzlich schon in einem früheren Durchlauf in mindestens einem Shift als sicher identifiziert worden sein.

In den Qualitätswert gehen folgende Parameter mit ein:



- Abstand ( $\Delta S$ ) der Shifts (S1, S2, evtl. S3) des experimentellen Peaks von der Zuordnung der entsprechenden Atome in der Datenbasis.
- Differenz der Abstände der entsprechenden Atome in der Struktur von den dem Peak entsprechenden Abstand im experimentellen Spektrum.
- Wenn eine Rückrechnung verwendet wird, statt der Differenz der Abstände die Differenz von simulierten und experimentellen Volumen<sup>-1/6</sup> (Pseudo- Abstände).

Der Qualitätswert gehorcht einer quadratischen Potentialfunktion die anhand der Parameter QualDist, QualW1 und QualW2 dafür sorgt, dass Peaks die sich um mehr als SmallLimit in einer Koordinate, oder um mehr als QualDistCutoff in ihrem strukturellen Abstand zum Abstand, ermittelt aus dem Peakvolumen, unterscheiden, nicht mehr als Erklärungsmöglichkeit betrachtet werden.

$Qual = \sqrt{a \cdot (\Delta shift_1)^2 + b \cdot (\Delta shift_2)^2 + c \cdot (\Delta dist)^2}$ , für den Fall, dass keine Rückrechnung verwendet wird.

$Qual = \sqrt{a \cdot (\Delta shift_1)^2 + b \cdot (\Delta shift_2)^2 + (\Delta d)^2}$ , für den Fall, dass ein Ergebnis- File einer Rückrechnung eingelesen wurde.

$$\Delta shift_1 = shift_{1exp} - shift_{1zuordnung}$$

$$\Delta shift_2 = shift_{2exp} - shift_{2zuordnung}$$

$$\Delta dist = dist_{exp} - dist_{struktur}$$

$$\Delta d = (Vol_{exp})^{-\frac{1}{6}} - (Vol_{sim})^{-\frac{1}{6}}$$

$a = QualW1$ ,  $b = QualW2$ ,  $c = QualDist :=$  Vorfaktoren, die dafür sorgen, dass der Qualitätswert >1 wird, wenn einer der Summanden ein zu schlechtes Ergebnis aufweist (vgl. Beschreibung der Parameter, Abschnitt 4.2.4).

- Qualitätswert > 1 wenn Peak von betrachteter Möglichkeit nicht im Rahmen der Zuordnung und der Eingabeparameter erklärt werden kann.
- Qualitätswert < 1 wenn Peak von betrachteter Möglichkeit erklärt werden kann.
- Qualitätswert = 0 wenn Peak ideal von Zuordnungsmöglichkeit erklärt wird.

#### **Anmerkung:**

Für die Ausgabe im textbasierten Zuordnungs- File nach dem Lauf von AUNOAS wird der Qualitätswert in die AUREMOL- übliche Qualitätswertgestaltung umgerechnet.

D.h. 1- interner Qualitätswert, sodass gilt:

- Qualitätswert = 0 wenn Peak von betrachteter Möglichkeit nicht im Rahmen der Zuordnung und der Eingabeparameter erklärt werden kann.

- Qualitätswert  $< 1$  wenn Peak von betrachteter Möglichkeit erklärt werden kann.
- Qualitätswert = 1 wenn Peak ideal von Zuordnungsmöglichkeit erklärt wird.

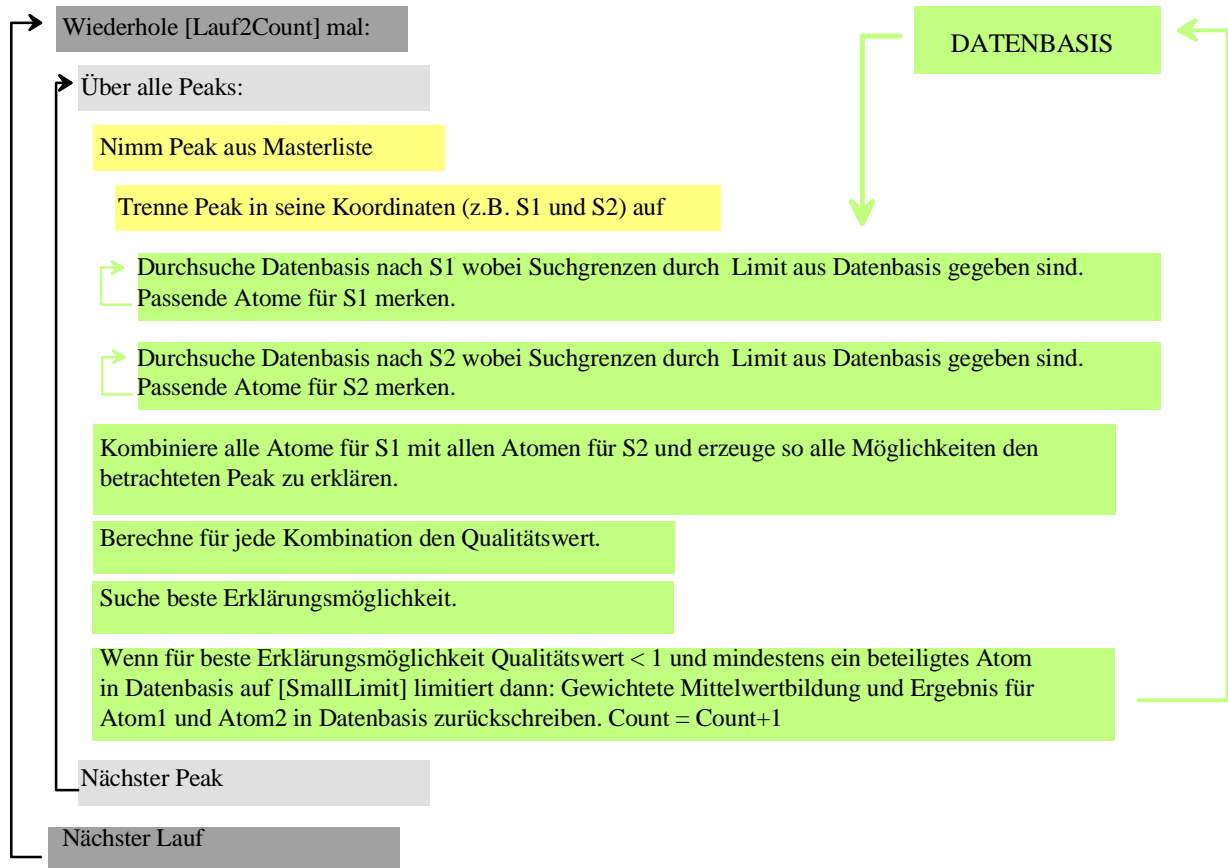
**Programmablauf:**

Abbildung 54: Schematische Darstellung des Programmteils [2].

**4.2.3.8. Verbleibende Peaks bearbeiten****Programmteil [3]**

Im letzten Schritt des Benennungsvorganges werden wiederum alle Peaks in [Lauf3Count] Durchläufen überprüft.

Einem experimentellen Peak wird die Bezeichnung zugeordnet, die den besten Qualitätswert ( $< 1$ ) aufweist.

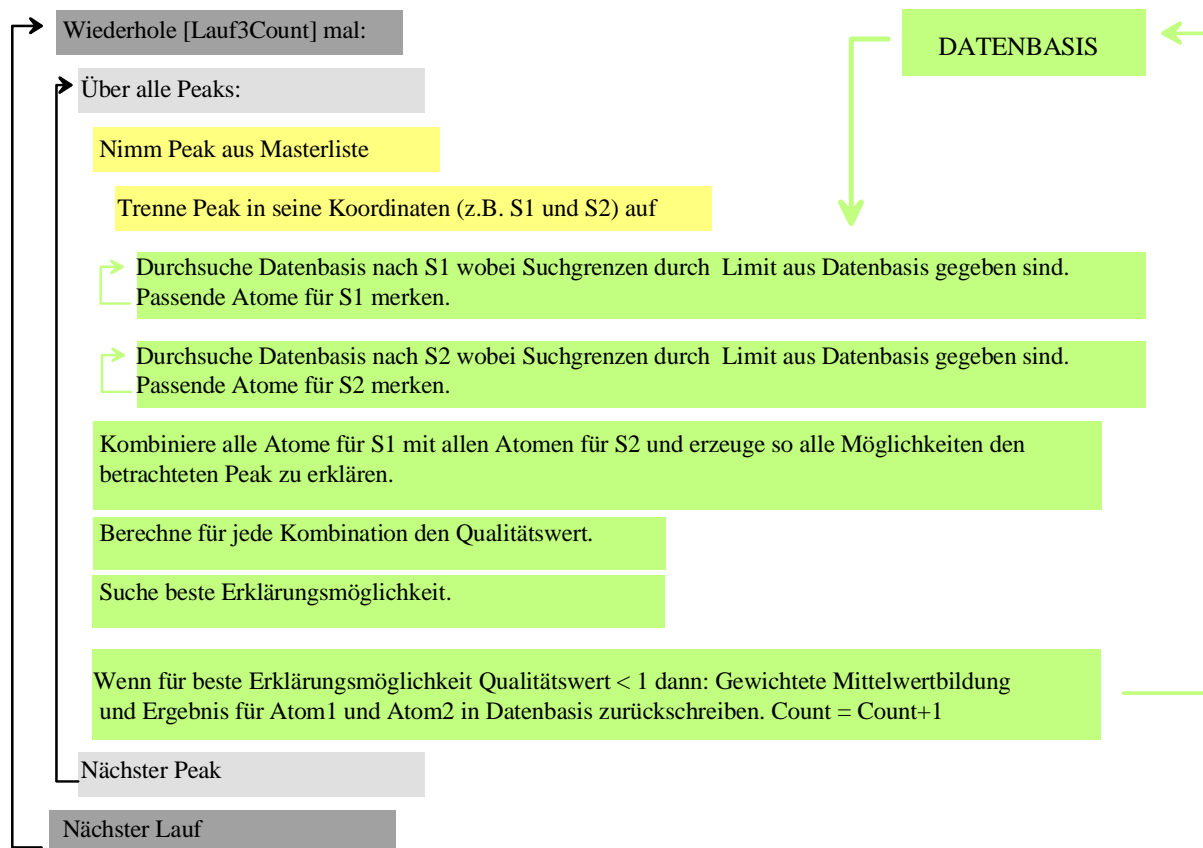
**Programmablauf:**

Abbildung 55: Schematische Darstellung des Programmteils [3].

**4.2.4. Programmparameter****Vom Benutzer einzugebende Parameter:**

BigLimit:	0,06 [ppm]
-----------	------------

Limit am Beginn des Suchvorganges.

Am Anfang des Benennungsvorganges werden alle Komponenten (Shifts) der zuzuordnenden Peaks im Umkreis von  $\pm \text{BigLimit}$  mit den ursprünglichen Zuordnungen (aus dem spt- File) verglichen. Eine Identifizierung des Peaks kann nur erfolgen, wenn ursprüngliche Zuordnung und Position des Peaks im Spektrum um nicht mehr als BigLimit voneinander abweichen.

Ein großer Wert [ca. 0,1 ppm] von BigLimit ist vorteilhaft, wenn von ungenauen Daten ausgegangen wird, das zuzuordnende Spektrum unter anderen Messbedingungen aufgenommen wurde, oder strukturelle Änderungen (z.B. Komplexbildung) untersucht werden sollen.

Ein kleiner Wert [ca. 0,02 ppm] ist anzuraten für hochaufgelöste Spektren und sehr präzise Anfangsdaten, um z.B. die Genauigkeit von Strukturen weiter zu verbessern.

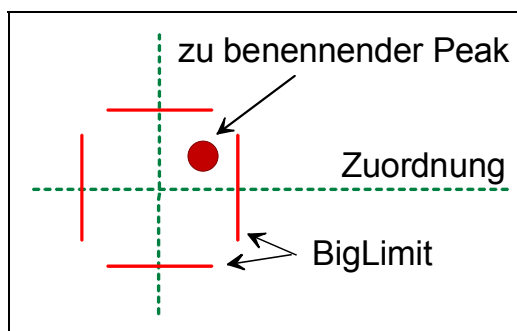


Abbildung 56: Illustration des Parameters „BigLimit“.

Generell ist zu sagen, dass die Werte für das Limit eines 3D Spektrums größer sein sollen, als für ein 2D Spektrum. Ein vernünftiger Ausgangspunkt für ein 2D Spektrum ist eine Kombination von Big- und SmallLimit von 0,06 und 0,03, für ein 3D Spektrum von 0,1 und 0,05. Je nach Größe und Beschaffenheit der Spektren müssen individuelle Einstellungen dieser Radien gefunden werden, um ein bestmögliches Ergebnis zu erzielen.

SmallLimit:	0,03 [ppm]
-------------	------------

Limit für zugeordnete Peaks (digitales Rauschen)

Im Verlauf des Benennungsvorganges passt das Programm anhand von sicher identifizierten Peaks die ursprüngliche Zuordnung immer weiter an das verwendete Spektrum an. D.h. das Kriterium, ab wann die Komponenten eines Peaks noch zu den Werten in der Zuordnung passen, wird immer strenger. Diesem Verhalten wird durch den Parameter SmallLimit Rechnung getragen. Atome deren Eintrag in der Datenbasis auf das SmallLimit gesetzt werden, sind im Verlauf des Benennungsprozesses als eindeutig im Spektrum identifiziert worden. Da ihr Shift aufgrund der Mittelwertbildung sich ebenfalls dem Wert im Spektrum angenähert hat, sind zu ihrer sicheren Identifizierung in anderen Peaks nun nur noch kleinere Toleranzen (SmallLimit) erlaubt.

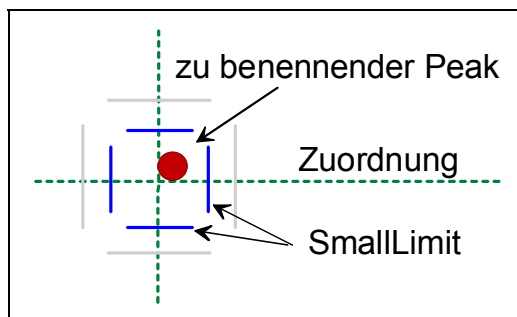


Abbildung 57: Illustration des Parameters „SmallLimit“.

Allerdings unterscheiden sich Peaks, die in einer Komponente zum gleichen Atom gehören, in dieser immer noch geringfügig. Dieser Unterschied entsteht hauptsächlich durch die begrenzte digitale Auflösung bei der Datenaufnahme. Für sehr gute Spektren sollte SmallLimit auf ca. 0,01 ppm gesetzt werden. Normalerweise bewährt sich aber ein Wert von  $0,5 \cdot \text{Biglimit}$ .

Scale1:	1
---------	---

Skalierungsfaktor für die erste indirekte Dimension bei der Setzung von Big- und SmallLimit (2D und 3D).

Scale2:	1
---------	---

Skalierungsfaktor für die zweite indirekte Dimension bei der Setzung von Big- und SmallLimit (3D).

Diese Skalierungsfaktoren definieren eine Verzerrung des Suchradius für die Indirekten Dimensionen. Dies ist dann sinnvoll, wenn man davon ausgehen kann, dass das verwendete Spektrum wesentlich geringer in den indirekten Dimension(en) aufgelöst ist, als in der direkten.

Die oben definierten Limits werden mit den angegebenen Skalierungsfaktoren multipliziert.

DistEindeutig	5,00
---------------	------

Distanz in der Struktur, ab der ein Peak noch als eindeutig gewertet wird.

Der Parameter DistEindeutig bestimmt, welchen Abstand die an einem Peak beteiligten Atome in der Struktur nicht überschreiten dürfen, um im Programmteil [1] noch als eindeutig identifiziert werden zu können. D.h. nur Signale, deren räumliche Distanz in der Struktur kleiner oder gleich DistEindeutig ist, werden als Erklärungsmöglichkeit in Betracht gezogen. Am Anfang der Strukturbestimmung, wenn noch keine vernünftige Strukturannahme besteht, sollte dieser Wert auf ca. 3-4 Angström gesetzt werden. Dies führt dazu, dass bevorzugt sequentielle und intraresiduelle NOEs berücksichtigt werden.

Liegt bereits eine brauchbare Strukturvorstellung vor, sollte DistEindeutig auf ca. 4,5 bis 5 Angström gesetzt werden, 6 Angström aber nicht überschreiten, was der "sichtbaren" Distanz im Spektrum entspricht.

#### ***Im Programm fest vorgegebene Parameter:***

Startcount:	10
-------------	----

Faktor für die Gewichtung der sequenziellen Zuordnung (spt/meta- File) am Anfang.

Für jedes Atom des Proteins existiert in der AUNOAS Datenbasis ein Feld für die Anzahl, wie oft dieses Atom im Benennungsprozess als eindeutig identifiziert wurde. Dieser Wert wird für die Mittelwertbildung verwendet. Diese Mittelung sorgt dafür, dass, sollte evtl. ein Atom einmal falsch zugeordnet werden; die Änderung des Shifts in der Datenbasis nicht allzu stark ausfällt. Deswegen ist es sinnvoll, den Wert der ursprünglichen Zuordnung am Anfang des Benennungsprozesses stärker zu gewichten um so Fehlzuordnungen zu verhindern. Je höher also der Wert von Startcount ist, desto langsamer (Dämpfung) passt sich die

ursprüngliche Zuordnung dem experimentellen Spektrum an. Für zu hohe Werte (ca. 30) ist es unter Umständen nicht mehr möglich, eine komplette Anpassung zu erreichen. Andererseits führen zu geringe Werte (ca. 1) von Startcount zu einer höheren Fehleranfälligkeit; und damit unter Umständen zu Fehlbenennungen. Für Spektren mit ca. 10- 20 Signalen pro Aminosäure sind Werte von 10 optimal.

Dieser Wert wird bei jeder Identifizierung des entsprechenden Atoms hochgezählt. (Vgl. Abschnitt 4.2.3.4, Berechnung des Mittelwertes).

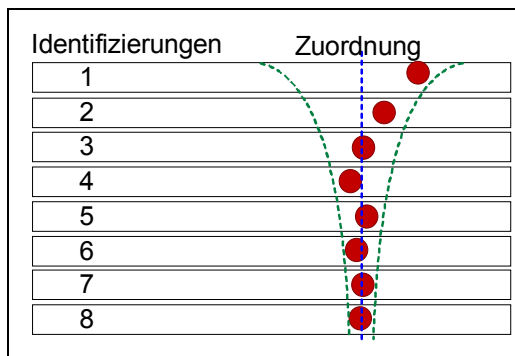


Abbildung 58: Schematische Darstellung der Anpassung der Zuordnung während des Benennungsprozesses.

ExpMinDist:	1,80
-------------	------

Minimaler Abstand der einem Peak im Experiment entsprechen kann.

ExpMaxDist:	5,00
-------------	------

Maximaler Abstand der einem Peak im Experiment entsprechen kann.

Während des Programmlaufes wird ein Algorithmus durchgeführt, der anhand der experimentellen Volumen und der bis dahin aufgrund ihrer exponierten Lage eindeutig identifizierten Peaks eine Umrechnung der Peakvolumen auf Distanzen in der Struktur durchführt. Um diese Eichung vornehmen zu können, muss dem Programm mitgeteilt werden, welchem strukturellem Abstand der intensivste (ExpMinDist) und der schwächste (ExpMaxDist) NOE im Spektrum entsprechen.

Normalerweise verwendet man für ExpMinDist 1,8 Angström und für ExpMaxDist 5,0 Angström.

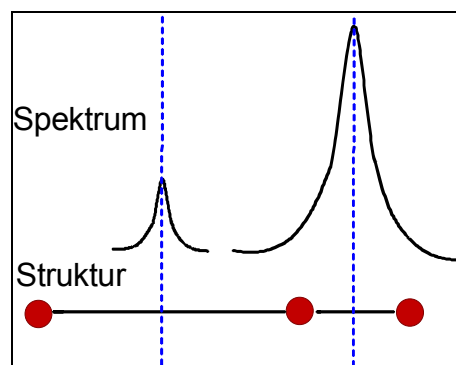


Abbildung 59: Je größer der Abstand in der Struktur, desto geringer die Signalintensität.

Lauf1Count:	20
-------------	----

Anzahl der Läufe im ersten Programmschritt, der eindeutige Zuordnungen findet.

Dieser Programmschritt wird so oft wiederholt, bis sich keine Änderungen mehr ergeben (Maximal 20).

Laufnochange	5
--------------	---

Anzahl der Läufe im Programmteil [1] ohne Änderung der Limits

Am Anfang des Benennungsvorganges versucht das Programm jeden Peak innerhalb von BigLimit anhand der sequenziellen Zuordnung zu benennen. Wenn es für diese Benennung nur eine einzige Möglichkeit gibt, so wird das Signal benannt, die entsprechenden Einträge in der Datenbasis werden gemittelt, und die Limitierungen werden von BigLimit auf SmallLimit gesetzt. Nur bei den ersten Durchläufen (Laufnochange) unterbleibt dieses Zurücksetzen der Limitierung, um den anhand der Mittelwertbildung korrigierten Shifts genug Zeit zu geben, sich um den dem experimentellen Spektrum entsprechenden Wert zu stabilisieren. In der Regel sind für den Parameter Laufnochange 5 Durchläufe ausreichend.

Insgesamt wird der erste Programmteil, der zur Identifizierung eindeutiger NOEs dient Lauf1Count-mal durchgeführt. In der Regel sind 20 Durchläufe mehr als genug, damit sich die eindeutig identifizierten Shifts gut an ihren neuen Wert annähern.

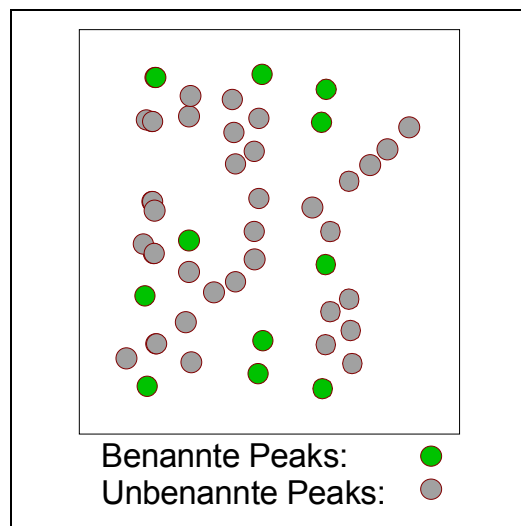


Abbildung 60: Zugeordnete isoliert liegende Signale nach dem ersten Programmschritt.

Lauf2Count	20
------------	----

Anzahl der Läufe für Programmschritt [2]. Peaks die in einer Koordinate sicher identifiziert sind, und zusätzlich einen guten Qualitätswert aufweisen, werden zugeordnet.

Dieser Programmschritt wird so oft wiederholt, bis sich keine Änderungen mehr ergeben (Maximal 20).

Im zweiten Schritt des Benennungsvorganges werden wiederum alle Peaks überprüft. Allerdings werden zunächst nur Peaks zugeordnet, die sowohl einen guten Qualitätswert (zur Definition des Qualitätswertes siehe Abschnitt 4.2.3.7, S. 68 ff) aufweisen, als auch zusätzlich in vorhergehenden Durchläufen in mindestens einer Koordinate sicher identifiziert worden sind.

Dieser Programmteil ordnet ausgehend von bereits gewonnenem Wissen anhand der nun schon teilweise an das experimentelle Spektrum angepassten Zuordnung, so viele Peaks wie möglich zu, und verbessert die experimentspezifische Zuordnung weiter.

In der Regel sind nach 20 Durchläufen (Lauf2Count) alle auf diese Weise identifizierbaren Peaks zugeordnet und die Werte ihrer chemischen Verschiebung haben sich in der experimentspezifischen Zuordnung stabilisiert.

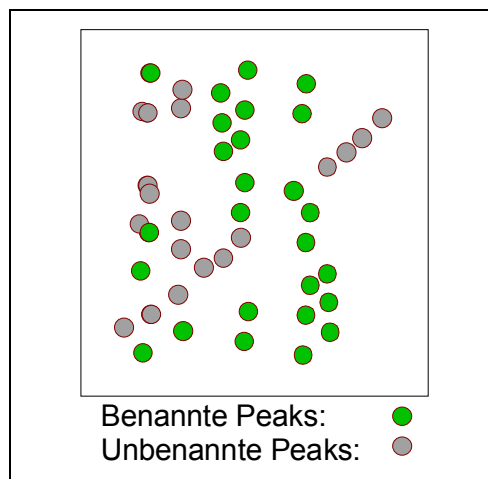


Abbildung 61: Zugeordnete in einer Koordinate bereits identifizierte Signale nach dem zweiten Programmschritt.

Lauf3Count	20
------------	----

Anzahl der Läufe für Programmschritt [3]. Hier werden alle Peaks anhand ihrer Qualität zugeordnet. Dieser Programmschritt wird so oft wiederholt, bis sich keine Änderungen mehr ergeben (Maximal 20).

Im abschließenden Schritt des Benennungsvorganges werden wiederum alle Peaks pro Durchlauf abgearbeitet. Da die experimentspezifische Zuordnung zu diesem Zeitpunkt schon sehr weit fortgeschritten ist, werden nun alle Peaks mit guten Qualitätswerten<sup>6</sup> zugeordnet, und ihre Komponenten gemittelt in die Zuordnung übernommen. Die Läufe (Lauf3Count) aus diesem Programmteil geben den noch vereinzelt neu identifizierten Shifts genügend Zeit um sich zu stabilisieren.

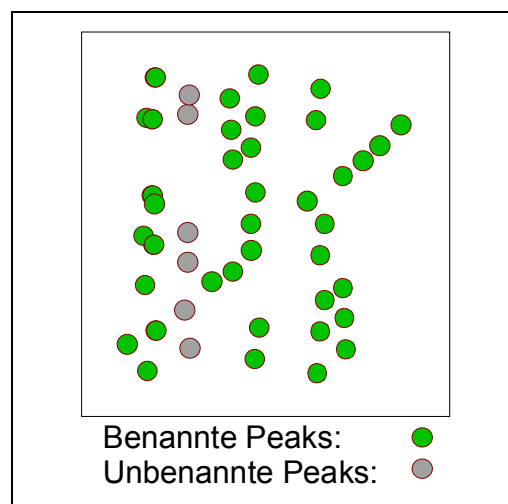


Abbildung 62: Benannte Signale nach dem dritten Programmschritt. Alle in der sequenziellen Zuordnung vorhandenen Spinsysteme sind identifiziert.

<sup>6</sup> Guter AUNOAS Qualitätswert für die Zuordnung: Qual << 1



DistQualität	6,00
--------------	------

Distanz in der Struktur, bis zu der maximal ein Peak noch erklärt wird.

In den Programmteilen [2] und [3] werden die Signale unter anderem anhand eines Qualitätswertes benannt, der beschreibt, wie gut der Peak zu der experimentellen Zuordnung passt. In diesen Qualitätswert geht unter anderem der Abstand der beteiligten Atome in der Struktur, sowie das Volumen im Spektrum (und damit der experimentell bestimmte Abstand) ein.

Der Parameter DistQualität bestimmt, ab welchem Abstand der beteiligten Atome in der Struktur der Qualitätswert soweit verschlechtert wird, dass eine Zuordnung der entsprechenden Atome zum aktuellen Peak nicht mehr durchgeführt wird.

Die Festlegung des Parameters auf 6 Angström geht davon aus, dass bei Nutzung des Programms eine brauchbare Strukturvorstellung vorliegt, und man so die für die endgültige Struktur notwendigen longrange NOEs finden kann.

QualDistCutoff	2,00
----------------	------

Parameter, der bestimmt, ab welcher Differenz zwischen struktureller Distanz und aus dem Spektrum ermittelter Distanz ein Peak noch gewertet wird.

Dieser Parameter bestimmt, ab welcher Differenz zwischen experimentell bestimmtem Abstand der Atome und ihrem strukturellen Abstand der Qualitätswert so schlecht wird, dass der entsprechende Peak nicht mehr zugeordnet wird.

QualDist	
----------	--

Funktionswert für die Formel in der Qualitätsberechnung. Wird berechnet anhand des oben beschriebenen Parameters.

$$[\text{QualDist}] = 1 / [\text{QualDistCutoff}]^2$$

QualW1	
--------	--

Funktionswert für die Formel in der Qualitätsberechnung. Wird berechnet anhand des eingegebenen Parameters.

$$[\text{QualW1}] = 1 / [\text{SmallLimit}]^2$$

QualW2	
--------	--

Funktionswert für die Formel in der Qualitätsberechnung. Wird berechnet anhand des eingegebenen Parameters.

$$[\text{QualW2}] = 1 / [\text{SmallLimit}]^2$$

Die hier aufgeführten Parameter sind wesentlich für die Funktion der Qualitätswertberechnung eines Peaks. Sie werden anhand der weiter oben beschriebenen Parameter automatisch berechnet. Aufgrund dieser Parameter wird in der Qualitätswertberechnung entschieden, wie gut ein Peak erklärt wird.

MasterQualitaet:	ja/nein
------------------	---------

Berücksichtigung des Qualitätswerts aus Masterliste für einen Peak.

Die eingelesene Masterliste kann ebenfalls einen Qualitätswert enthalten, der eine Aussage darüber trifft, mit welcher Wahrscheinlichkeit der betreffende Peak ein echtes NOE Signal ist, oder ob es sich eher um ein Rauschsignal handelt. Dieser Qualitätswert hat **nichts** mit dem oben erwähnten Qualitätswert für die Peakzuordnung zu tun.

Der Parameter MasterQualitaet bestimmt, ob der Wert aus der Masterliste vorhanden ist. Wenn er vorhanden ist, wird er verwendet, und sorgt bei der Mittelwertbildung dafür, dass „schlechte“ Peaks deutlich weniger gewichtet werden.

## 4.3. Ergebnisse

### 4.3.1. Ausgabe- Files

Wenn das Programm beendet wurde, informiert ein Fenster den Benutzer darüber, wie viele Signale zugeordnet wurden, und ob die Zuordnung erfolgreich war.



*Abbildung 63: Ende des Programmlaufes. Information über die Anzahl der zugeordneten Signale.*

Die Masterliste wird automatisch aktualisiert, und der Benutzer kann das Ergebnis direkt am Bildschirm anhand der jetzt benannten Signale betrachten.

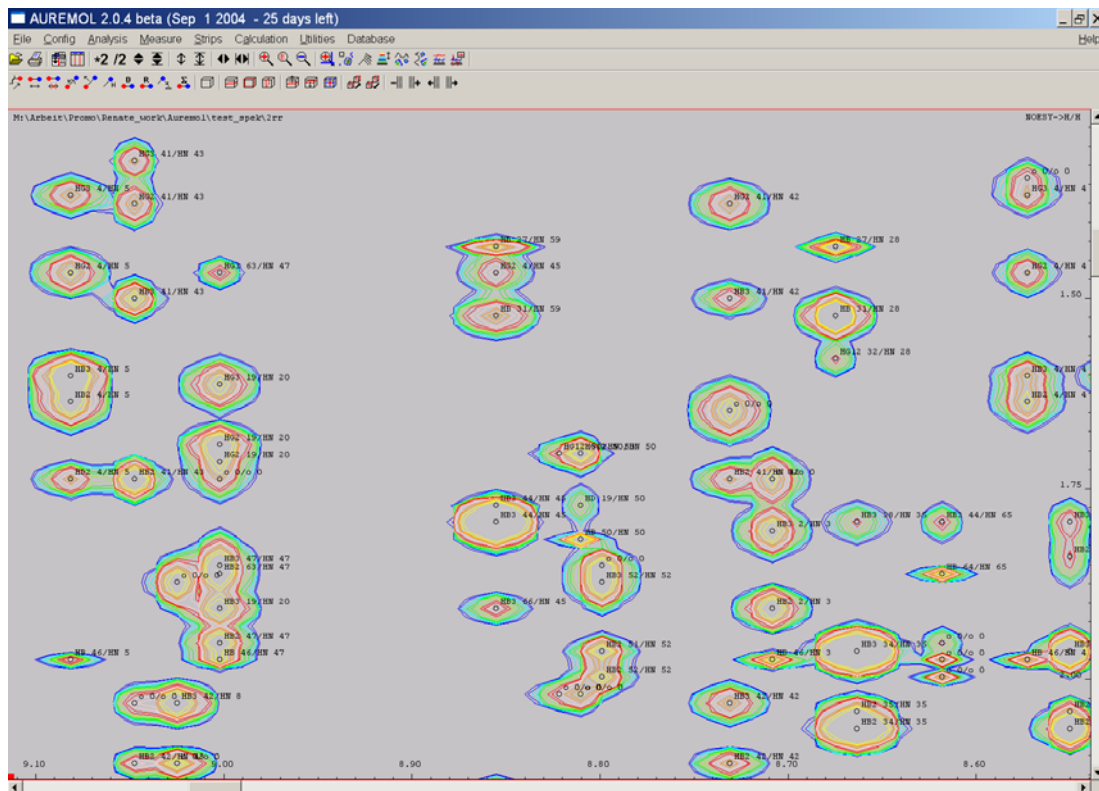


Abbildung 64: Screenshot von AUREMOL nach Beendigung des Programmlaufs.

Zusätzlich zur Aktualisierung der Masterliste nach dem Programmlauf, werden zwei textbasierte Ausgabe-Files erzeugt, teilweise mit zusätzlichen Informationen. Diese Ausgabe-Files können in anderen Programmen, beispielsweise EXCEL weiterverwendet werden.

Das **output- File assignment** enthält die Ergebnisse der Zuordnung, die im Programm berechnete Distanz, das Volumen, und den intern berechneten Qualitätswert.

Das **output- File shifts** enthält die an das aktuelle Spektrum angepasste (veränderte) sequenzielle Zuordnung, in einem Format, das dazu benutzt werden kann, mit AUREMOL ein neues **meta- File** zu erzeugen.

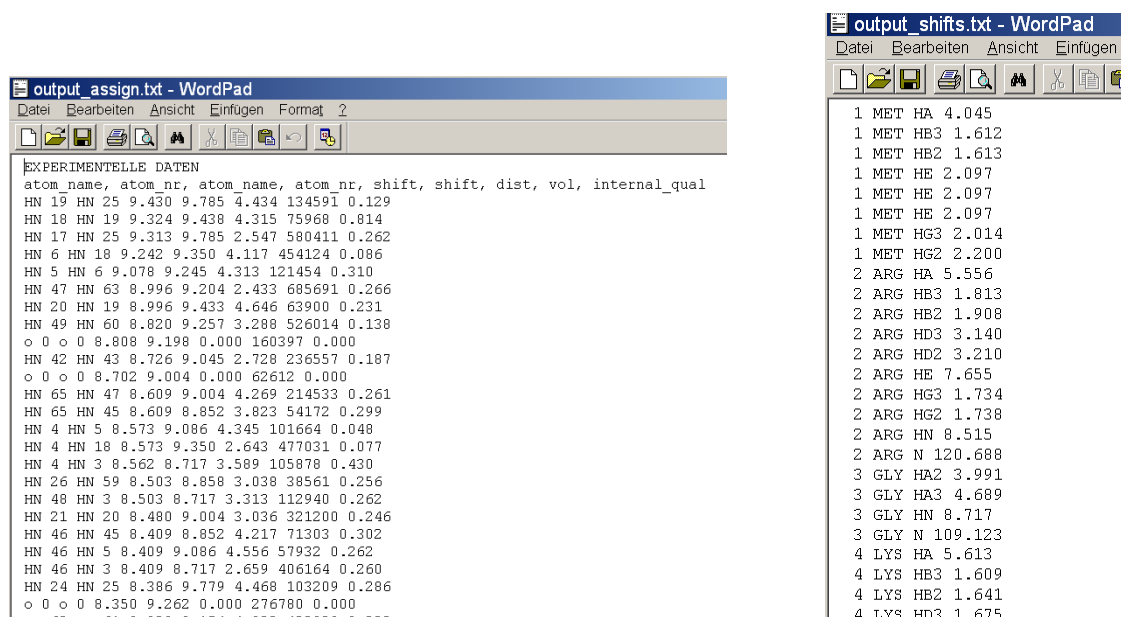


Abbildung 65: Ausgabedateien des Programms AUNOAS. ASCII- Darstellung der Ergebnisse.

### 4.3.2. Testergebnisse

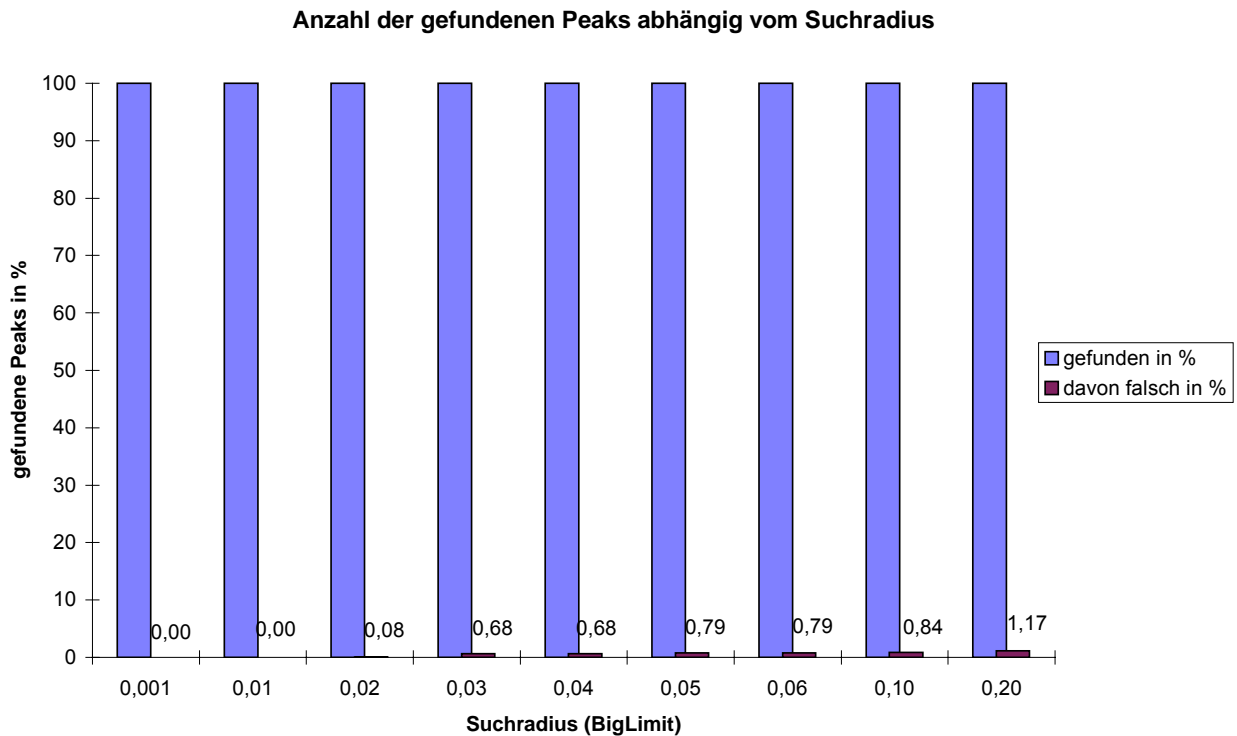
#### **Test des Programms mit synthetischen Daten.**

Zur Überprüfung der richtigen Funktionsweise haben wir das Programm zunächst mit synthetischen, überprüfbaren Daten getestet.

Ein Spektrum des Kälteschockproteins *TmCSP* wurde mittels pdb- und spt- File analog der Simulation in Abschnitt 3.3.2.2 rückgerechnet. Das ergab eine Masterliste mit ca. 6000 benannten Peaks. Diese Masterliste wurde AUNOAS als experimentelle, unbenannte Masterliste übergeben und zugeordnet.

Nach dem Programmlauf wurde überprüft, wie viele Signale richtig benannt worden waren. Das Diagramm stellt dar, wie viele der vorhandenen Signale (in %) zugeordnet wurden, und wie hoch die jeweilige Fehlerquote bei unterschiedlichen verwendeten Radien für BigLimit ist.

Bei einem Spektrum mit 6000 Signalen liegen diese sehr dicht, außerdem ist eine große Zahl an überlappenden Peaks zu erwarten, weswegen ein Wert von 0,2 ppm für BigLimit, der auch für experimentelle Spektren mit ca. 1000 Signalen schon sehr hoch ist (vgl. oben, empfohlener Wert für 2D NOESYs= 0,06 ppm), einen ziemlichen Härtestest für das Programm darstellt. Trotzdem bleibt auch für diesen hohen Fangradius die Fehlerquote ausnehmend gering (ca. 1%).



*Abbildung 66: Test des Programms mit synthetischen Daten. Die blauen Balken zeigen wie viele (in %) der vorhandenen 6000 Signale zugeordnet wurden, die roten Balken zeigen an, wie viele davon falsch sind.*

Es zeigt sich, dass das Programm eine sehr hohe Zuordnungswahrscheinlichkeit (immer 100%) bei einer nahezu vernachlässigbaren Fehlerquote aufweist. Selbst vergleichsweise „riesige“ Fangradialen von 0,2 ppm bei einem Spektrum mit 6000 Peaks führen zu einem guten Ergebnis.

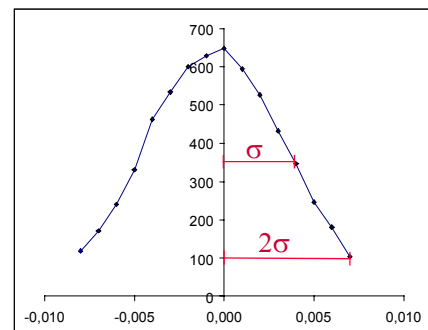
Dieser erste Test illustriert die zuverlässige Funktionsweise des verwendeten Algorithmus.

#### **Test des Programms mit verrauschten synthetischen Daten.**

Daran anschließend wurde die oben verwendete synthetische Masterliste des *TmCSP* Proteins von uns in der w1 und w2 Koordinate mit den jeweils angegebenen Werten verrauscht (Verschieben der Position der einzelnen Signale).

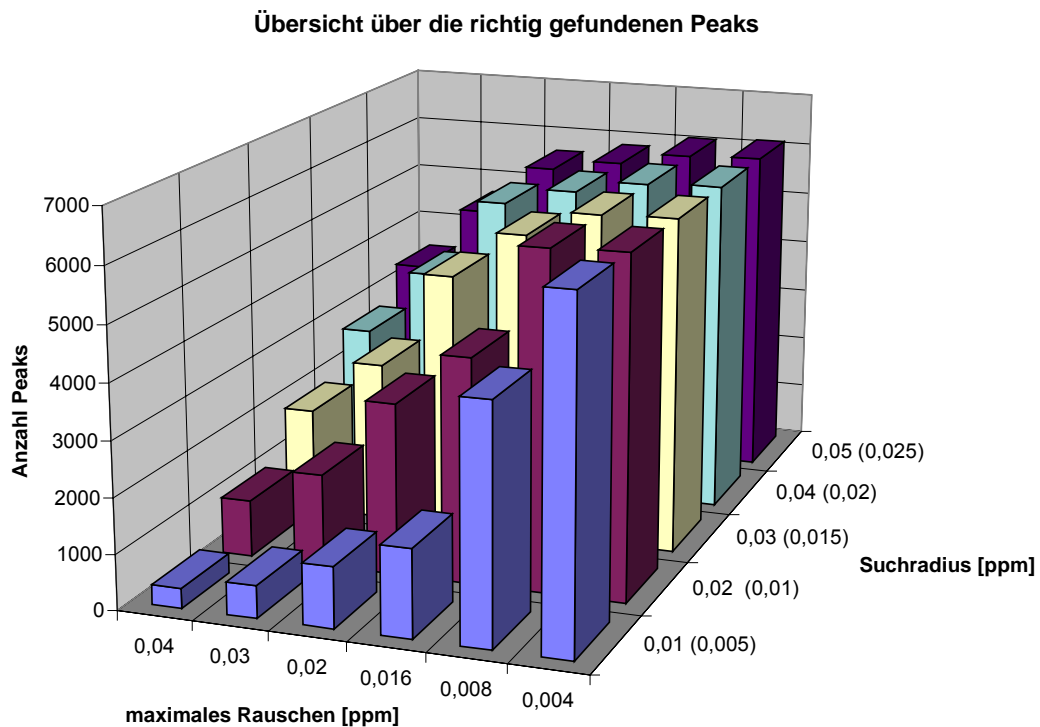
Dieser Test simuliert ein reelles Spektrum, bei dem aufgrund verschiedener Faktoren nicht alle Signale eines Spinsystems exakt auf einer Linie liegen.

Das Rauschen wurde anhand einer Gauß-Verteilung mittels Zufallszahlen erzeugt, wobei das maximale, im Diagramm angegebene Rauschen  $2 \cdot \sigma$  entspricht. Getestet wurden wiederum verschiedene Fangradialen.



*Abbildung 67: Beispiel für eine verwendete Rauschfunktion ( $\sigma=0,004$ ).*

BigLimit und SmallLimit (in Klammern) angegeben.

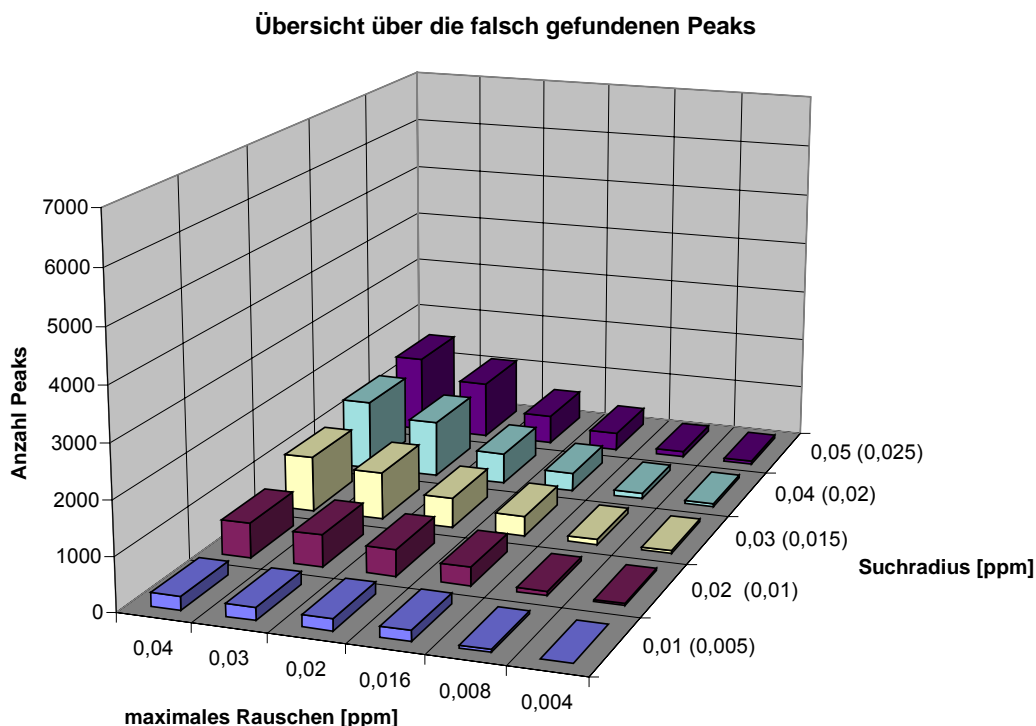


*Abbildung 68: Darstellung der richtig zugeordneten Signale bei unterschiedlichen Kombinationen aus Rauschen und Fangradius.*

Das Diagramm zeigt, dass das Programm unter realistischen Bedingungen (wenn der kleinere Fangradius (SmallLimit), der die Ungenauigkeit der Position der Peaks ausgleicht, größer als das auf den Peaks liegende Rauschen ist) exakt zuordnet (ca. 100% der Peaks zugeordnet). Natürlich ist es für Fälle, in denen das Rauschen auf dem Spektrum den Fangradius übersteigt, auch dem Programm nicht mehr möglich Erklärungen für die Signale zu finden.

Gezeigt hat sich bei den Tests, dass es für unterschiedlich verrauschte Spektren bestmöglich zusammenpassende Kombinationen aus Rauschen und Fangradius gibt. Für das beste Zuordnungsergebnis sollte immer der kleinstmögliche Fangradius (SmallLimit) gewählt werden, der das Rauschen gerade noch abdeckt.

Dementsprechend ist über einen zusätzlichen Programmteil nachzudenken, der automatisch die Qualität eines Spektrums erkennt, und den dafür zu wählenden Fangradius automatisch anpasst.



*Abbildung 69: Test des Programms mit verrauschten synthetischen Daten. Die synthetische Masterliste des TmCSP Proteins wurde wie oben angegeben verrauscht und zugeordnet. Dargestellt ist hier eine Übersicht darüber, wie viele der Signale falsch zugeordnet wurden, aufgetragen anhand Rauschen und Fangradius.*

Wie man auch hier sieht, ist die Fehlerquote sehr gering für Fälle, in denen das Rauschen den Fangradius nicht übersteigt. Weiterhin ist zu erkennen, dass auch bei Fällen, in denen die Aufgabe des automatischen Zuordnens unrealistisch wird, das Programm hauptsächlich mit der Nichtzuordnung der betreffenden Signale reagiert, nicht aber übermäßig viele Fehler macht. Diese Tatsache ist ein nicht zu unterschätzender Vorteil bei der Zuordnung experimenteller Spektren, da bei diesen ein falsch zugeordnetes Signal schwerer wiegt, als viele unzugeordnete.

### 4.3.3. Stabilität

Um auch andere, in experimentellen Spektren auftretende Störungen und ihre Auswirkungen auf das Programm zu untersuchen, haben wir unterschiedliche Störungen simuliert.

Zunächst der „Störungstyp“ der Verunreinigung, was bedeutet, dass neben dem zu messenden Proteinsignalen gleichwertige Signale existieren, die z.B. von teilweise denaturiertem Protein oder von vorhandenem Fremdproteinen oder anderen Zusätzen stammen können. Um dieses zu untersuchen, wurden Signale hinzugefügt, die gleich hohe

Volumen wie die eigentlichen Signale aufwiesen, und auch einen guten Qualitätswert von AUREMOL zugewiesen bekamen, da diese Signale ja auch wirkliche Signale sind.

Als zweites wurden Störungen hinzugefügt, die durch Artefakte zustande kommen. Artefakte entstehen durch die Aufnahme und Prozessierung von Spektren und weisen gleiche Volumen wie die Signale auf, werden aber aufgrund ihrer eigenen anderen Peakform und Symmetrie von AUREMOL mit niedrigen Qualitätswerten bedacht.

Zuletzt sollte noch Rauschen simuliert werden, das in experimentellen Spektren allgegenwärtig ist, und den manuellen Zuordnungsprozess sehr erschwert. Rauschpeaks zeichnen sich durch ein niedrigeres Volumen und damit verbunden auch durch einen niedrigeren Qualitätswert aus. Allerdings können auch wirkliche Signale stark verringerte Volumina aufweisen, wenn diese Signale von sehr langreichweitigen Kontakten stammen. Dann wiederum sind diese Peaks von großer Bedeutung für die spätere Strukturbestimmung. Deshalb ist es sehr wichtig, dass ein Programm zur automatischen Zuordnung von Spektren auch möglichst genau zwischen Rauschen und Signal unterscheidet.

Qualitätswert der Störungen	Volumen zu Peakvolumen	Störungstyp	Anzahl der CSP Peaks	Anzahl der Rauschpeaks	Zugeordnet CSP	CSP falsch zugeordnet	Rauschpeaks zugeordnet
1	1 zu 1	Verunreinigungen	6168	100	6168	0	1
1	1 zu 1	Verunreinigungen	6168	500	6168	0	3
1	1 zu 1	Verunreinigungen	6168	1000	6168	0	6
1	1 zu 1	Verunreinigungen	6168	2000	6168	0	8
1	1 zu 1	Verunreinigungen	6168	4000	6168	0	17
0,1	1 zu 1	Artefakte	6168	100	6168	0	1
0,1	1 zu 1	Artefakte	6168	500	6168	0	0
0,1	1 zu 1	Artefakte	6168	1000	6168	0	6
0,1	1 zu 1	Artefakte	6168	2000	6168	0	6
0,1	1 zu 1	Artefakte	6168	4000	6168	0	15
0,1	1 zu 0,1	Rauschen	6168	100	6168	0	1
0,1	1 zu 0,1	Rauschen	6168	500	6168	0	5
0,1	1 zu 0,1	Rauschen	6168	1000	6168	0	3
0,1	1 zu 0,1	Rauschen	6168	2000	6168	0	9
0,1	1 zu 0,1	Rauschen	6168	4000	6168	0	18

*Abbildung 70: Darstellung der untersuchten Störungen und der Reaktion des Programms. Es ist zu erkennen, dass in keinem der untersuchten Fälle das Programm ein Proteinsignal falsch zugeordnet hat.*

Wie Abbildung 70 zu entnehmen ist, reagiert das Programm stabil auf jeden angebotenen Störungstyp. Egal, ob es sich um Verunreinigungen, Artefakte oder Rauschen handelt, und auch weitgehend unabhängig von der absoluten Anzahl der vorhandenen Störungen (bis hin zu nahezu 1:1 Verhältnis von wirklichen Signalen und Störung!). In allen Fällen wurden die Signale des verwendeten Proteins richtig und vollständig zugeordnet. Mit zunehmender Anzahl der Störungen wurden bis maximal 18 Signale, die Störungen darstellen, zugeordnet. Eine verschwindend geringe Zahl bei einem Signalaufkommen von dabei mehr als 10000 Peaks.



### 4.3.4. Anwendungsbeispiele

Wir haben das Programm AUNOAS entwickelt, um die zeitaufwendige Benennung von NOESY Spektren zu automatisieren. Somit bieten sich als Einsatzgebiete alle Vorgänge an, die bisher nur unter großem Zeitaufwand zu bewerkstelligen waren.

#### **Qualitätssicherung**

Da NOESY Spektren viele für die Proteinstruktur relevanten Informationen enthalten, bietet sich ihr Einsatz z.B. für die Überwachung der Produktion im pharmakologischen Bereich an. Um dieses Vorhaben zu testen, haben wir ein synthetisches Spektrum aus den Spektren von Saratin (8047 NOEs) und CSP (6168 NOEs) generiert. Dazu wurden die beiden rückgerechneten 2D NOESYs in einer gemeinsamen Masterliste zusammengefasst und gemischt. Außerdem wurde ein pdb- und ein spt- File mit den Daten aus beiden Proteinen erzeugt.

Das Programm war in der Lage die Signale von beiden Proteinen zu trennen und richtig zu benennen. Damit erscheint es z.B. möglich, einfache Proteingemische zu untersuchen, oder zu entscheiden, wenn nur ein Protein in der Lösung enthalten ist, wie viel Prozent richtig, und wie viel falsch gefaltet sind.

#### **Temperatur-, Druck-, pH- Reihen**

Als einer der ersten Praxistests haben wir mit AUNOAS eine Temperaturreihe für Saratin erstellt. Dazu wurden ausgehend von 298K (für diese Temperatur existierte eine vollständige Zuordnung) 2D NOESY Spektren mit einer jeweils um  $\pm 1$ K veränderten Temperatur zugeordnet. Ausgehend von diesen Zuordnungen wurden dann wieder neue Spektren mit wieder um  $\pm 1$ K veränderten Temperaturen gemessen und zugeordnet.

Auf diese Ergebnisse wird im Einzelnen im nächsten Abschnitt ausführlich eingegangen.

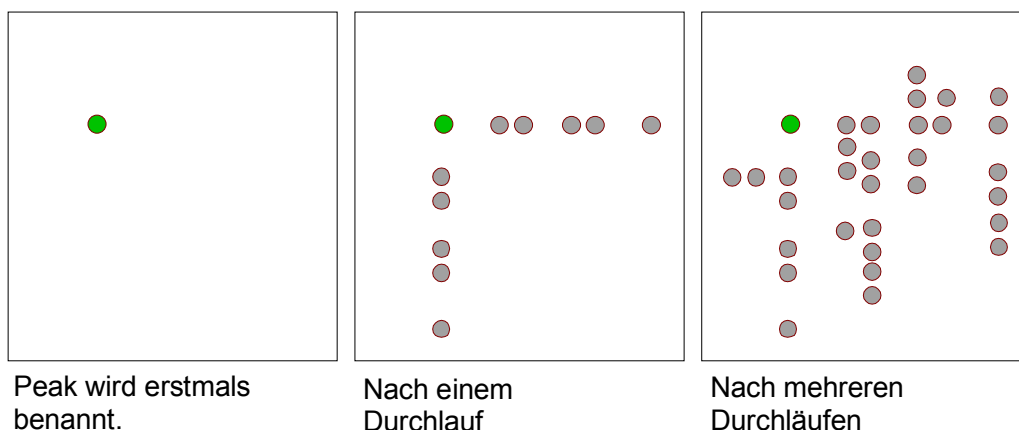
## 4.4. Zusammenfassung und Ausblick

### 4.4.1. Funktionalität

Betrachtet man eine Signal, das während der Programmausführung benannt wird, so stellt dies auf den ersten Blick einen isolierten Prozess dar. Es sieht so aus als ob das Spinsystem, zu dem das Signal gehört, nicht berücksichtigt wird.

Da aber jeder Peak der benannt wird, die Shifts der beteiligten Atome in der Datenbasis beeinflusst, zeigt sich, dass bereits während eines Durchlaufes diese Beeinflussung alle Signale eines Spinsystems betrifft. Den Peak der benannt wurde, betrifft diese Änderung der Datenbasis erst im nächsten Durchlauf wieder. So gesehen, ist bereits einen Durchlauf später das betrachtete Signal mit allen anderen Signalen, die ein Atom mit ihm gemeinsam haben verbunden (gekoppelt). Nach einigen weiteren Durchläufen baut sich diese Kopplung

(anfangs schwach, dann immer stärker werdend) zwischen allen Peaks und allen Atomen des Proteins auf.



*Abbildung 71: Schematische Darstellung wie die Anzahl der einen Peak beeinflussenden anderen Signale abhängig von der Anzahl der Durchläufe zunimmt.*

Die Mittelwertbildung und die Tatsache, dass jedes Atom in dem betrachteten Spinsystem immer die gleiche chemische Verschiebung ( $\pm$  gewisse Ungenauigkeiten) haben muss, sorgen nun dafür, dass sich Spinsysteme bilden, und dass sich auch überlagerte Spinsysteme trennen können. Weiter werden die anfangs möglichen Fehlzuordnungen immer ungünstiger bewertet, da sie ja zu anderen Signalen nicht kompatibel sind. Diese Fehlzuordnungen werden im Laufe des Programms wieder eliminiert. Dies führt nach vielen Läufen dazu, dass eine neue sequenzielle Zuordnung generiert wird, die das zu benennende Spektrum unter Verwendung der anfänglichen sequenziellen Zuordnung so gut wie möglich erklärt.

#### 4.4.2. Verbesserungsmöglichkeiten und Ausblick

AUNOAS verwendet für die Benennung eines Spektrums einen Standardparametersatz, der vom Benutzer an seine Bedürfnisse angepasst werden kann. Dieser Einstellungsprozess verbessert bei nicht alltäglichen Anwendungen das Ergebnis, verlangt aber eine gewisse Erfahrung in der Wahl der Parameter.

Wünschenswert wäre ein vorgeschalteter Programmteil, der die verwendeten Daten analysiert und danach automatisch die Eingabeparameter setzt.

Eventuell würde auch eine Anpassung der Programmparameter während des Programmlaufes zu einer weiteren Verbesserung der Ergebnisse beitragen.

Momentan ist das Programm AUNOAS primär dazu bestimmt, NOESY- Peaks in NOESY-Spektren zu benennen. Durch geeignete Wahl der Steuerparameter ist es aber auch möglich, Strukturen zu verbessern. Allerdings laufen diese Prozesse immer noch zum Teil in

Handarbeit ab, da Restraint- Files erzeugt werden müssen und die Ergebnisse der Strukturrechnungen wieder für AUNOAS verfügbar gemacht werden müssen.

Hier wäre ein Ansatzpunkt den Strukturbestimmungsprozess weiter zu automatisieren.

Auch die Aufgabe der Beschränkung auf NOESY Spektren könnte sich als sinnvoll erweisen. Allerdings wären hierfür umfangreiche Änderungen in der Programmstruktur notwendig. Eventuell wäre es hierfür zweckmäßiger, für verschiedene Spektrentypen verschiedene Auswertprogramme zu verwenden, die allerdings auf eine gemeinsame Datenbasis zugreifen.

Für die Spezialbehandlung von nicht benennbaren Peaks haben wir schon Überlegungen angestellt, die eine weitergehende Anwendbarkeit des Programms gewährleisten würde.

Ausgehend von gut gewählten Eingabeparametern, einer guten Struktur und einem guten Spektrum, welches benannt werden soll, gibt es verschiedene Erklärungsmöglichkeiten für Peaks, die nicht benannt worden sind:

- Es handelt sich um einen Artefakt- oder Rausch- Peak.
- Die Ursprüngliche Zuordnung aus dem spt- File enthält einen Fehler oder ist unvollständig.
- Die gemessene Probe weicht teilweise von der, der Zuordnung zugrundeliegenden, ab (z.B. Komplex, Temperatur, pH, Salzgehalt, etc.).

Wenn AUNOAS das experimentelle Spektrum zugeordnet hat, sollten, wenn obige Fälle auftreten, einige Peaks unbenannt bleiben. Soweit es sich dabei um Artefakte, bzw. um Rauschpeaks handelt, ist dies auch erwünscht.

Allerdings wäre es sehr nützlich, ausgehend von den vorhandenen Zuordnungen auch noch Spinsysteme erfassen zu können, die zwar im Spektrum auftauchen, aber nicht identifiziert werden können, da sie entweder nicht, oder falsch in der sequenziellen Zuordnung vorhanden sind.

Nach dem Ablauf von AUNOAS sollten alle Atome der internen Datenbasis im experimentellen Spektrum identifiziert, und ihre Shifts diesem angepasst worden sein. Trifft dies nicht zu, so besteht die Möglichkeit, dass das nicht identifizierte Atom an einem nicht zugeordneten Spinsystem beteiligt ist.

Ausgehend von dieser Überlegung sucht man nun dieses Atom in der Proteinstruktur und ermittelt alle strukturellen Nachbarn mit einem Abstand von weniger als 5 Angström.

Da aber viele dieser Nachbarn während des AUNOAS- Laufes identifiziert worden sind, und man außerdem ihren strukturellen Abstand kennt, kann man mit Hilfe der Volumen-Distanzumrechnung ein „imaginäres Spinsystem“ kreieren, das diese Kontakte aufweist. Nun durchsucht man alle noch nicht zugeordneten Peakssysteme, ob sie zu dem imaginären Spinsystem passen. Ist dies der Fall, so wurde der fehlende Atom- Shift ermittelt.

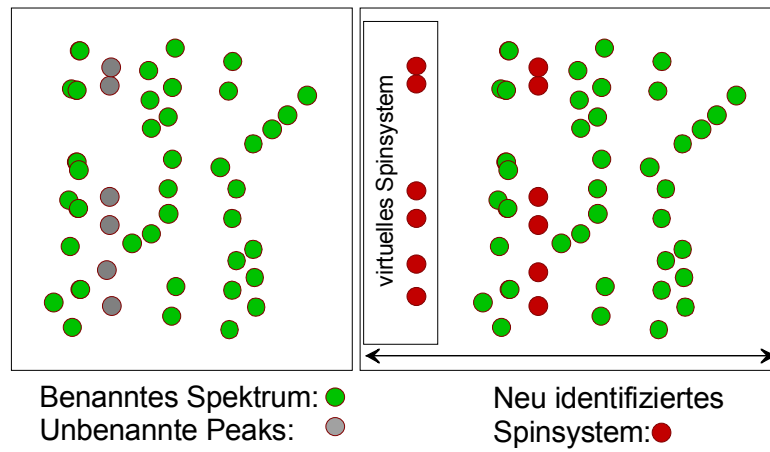


Abbildung 72: Bildliche Darstellung der Vorgehensweise bei der Identifikation nicht benannter Spinsysteme.

Dieses Verfahren sollte in der Lage sein von einer 80%igen Zuordnung die restlichen Signale zu identifizieren.

## 5. Temperaturreihe von Saratin

### 5.1. Daten

#### 5.1.1. Erzeugung

Ausgehend von den vorher beschriebenen Möglichkeiten zur automatischen Zuordnung von 2D NOESY- Spektren haben wir in einem umfangreichen Praxistest eine Temperaturreihe eines Proteins in Angriff genommen. Dazu wurde Saratin, ein Protein zur Inhibition der Blutgerinnung aus dem Blutegel, verwendet. Dieses Protein ist über einen weiten Temperaturbereich (Hitzenedenaturation bei ca. 75°C) stabil, liefert gute Signale, und war in ausreichender Menge am Lehrstuhl vorhanden. Außerdem existiert für Saratin eine nahezu vollständige sequenzielle Zuordnung und eine gute Vorstellung über die räumliche Struktur bei 298K [Dom04]. Saratin besteht aus 103 Aminosäureresten, die ein fünfsträngiges  $\beta$ -Faltblatt und eine  $\alpha$ - Helix bilden.

Mit diesem Protein haben wir in einer zweieinhalb Wochen dauernden, unterbrechungs-freien, Messreihe 41 2D- NOESY- Spektren in einem Temperaturbereich von 10°C bis 50°C in 1° Schritten aufgenommen. Die Zahl der gemessenen Punkte betrug in der w1- Dimension 1024, in der w2- Dimension 4096, und die Spektrenweite 14 ppm.

Die erzeugten Spektren wurden mit einer ungelabelten Probe von Saratin in 5 % D<sub>2</sub>O (v/v) bei pH 6,5 aufgenommen. Die Probe enthielt 40mg/ml Saratin, 10mM Na-P, 50mM NaCl, 0,2 mM DSS und 5 $\mu$ l einer Proteaseninhibitormischung. Gemessen wurde an einem Bruker DRX600 Spektrometer mit Kryo- Probenkopf (Protonenresonanz bei 600 MHz). Die chemischen Verschiebungen der Protonen wurden auf den internen Standard DSS (4,4-Dimethyl-4-Silapentan-Sulfonsäure) referenziert. Die Aufnahme und Prozessierung der Spektren erfolgte mit der Software XWINNMR 2.6 der Firma Bruker, Karlsruhe. Die Prozessierung umfasste neben der Fouriertransformation standardmäßig die richtige Phasierung und eine Basislinienkorrektur.

Die Probe von Saratin erwies sich während der ganzen Messdauer von zweieinhalb Wochen am Stück und dem großen Temperaturbereich von 283K bis 323K als sehr stabil. Nur aufgrund dieser Tatsachen konnte ein wirklich zusammenhängender Datensatz unter völlig identischen Randbedingungen erzeugt werden, der es, wie man später sieht, ermöglicht, Aussagen über das Temperaturverhalten von Saratin zu treffen.

Auch das Vorhandensein eines Kryo- Probenkopfes trug erheblich zur Güte der erhaltenen Spektren bei. Durch die nur einige Kelvin betragende Temperatur am Messkopf konnte das Signal- zu Rauschverhältnis signifikant verbessert werden, was in den Spektren deutlich sichtbar ist. Die aufgenommenen 2D NOESYs stellen sich hochaufgelöst und mit sehr großer Signalanzahl, verbunden mit niedrigem Rauschlevel dar.

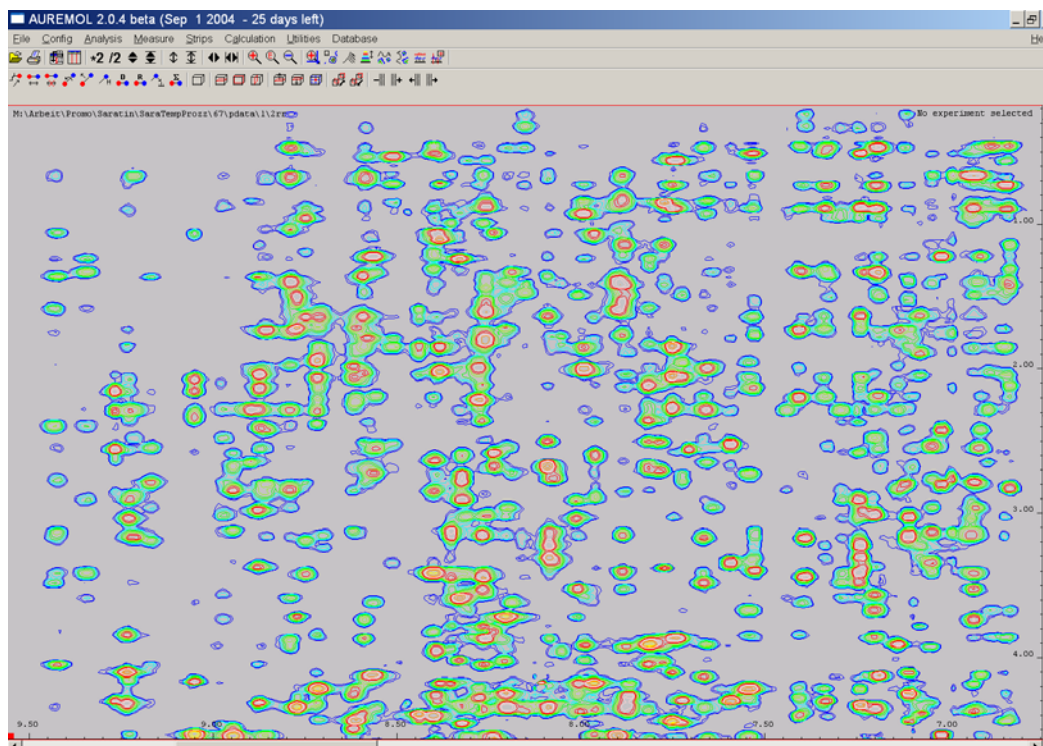


Abbildung 73: Ausschnitt aus einem 2D- NOESY Spektrum. Deutlich erkennbar sind die hohe Anzahl und die gute Auflösung der Signale.

Für die geplante automatische Zuordnung sind damit optimale Startbedingungen geschaffen worden, die das Arbeiten ohne menschliches Eingreifen ermöglichen.

### 5.1.2. Aufbereitung

Nach der Aufnahme und Prozessierung der Spektren wurden diese zunächst einem „Peakpicking“- Algorithmus unterzogen (siehe 2.2.1) und Signale aufgrund ihrer automatisch bestimmten Wahrscheinlichkeit, ein Protein- Peak zu sein, klassifiziert. Aufgrund der hohen Qualität und der kaum vorhandenen Rauschsignale der Spektren mussten jedoch alle Spektren anschließend von Hand noch einmal überprüft und gepickt werden, um auch schwache Signale zu identifizieren. Ziel war es, möglichst viele Signale zu erhalten, um auch langreichweitige Kontakte ausfindig zu machen. Eventuell gepickte Rauschsignale sollten aufgrund eingangs ausgeführter Testreihen und des verwendeten Algorithmus keinen störenden Einfluss auf den Ablauf des Programms haben.

Nach diesem Prozess waren pro Spektrum durchschnittlich 4000 bis 5000 Signale vorhanden. Diese wurden nun integriert und einer automatischen Volumenberechnung unterzogen.

Mit den so vorbereiteten Daten haben wir die automatische Zuordnung begonnen.

## 5.2. Automatische Zuordnung mit AUNOAS

Bei der automatischen Zuordnung der Spektren haben wir mit dem Spektrum für 298K begonnen, da für diese Temperatur die Originalstruktur von Saratin erzeugt wurde und dafür auch eine passende sequenzielle Zuordnung existierte.

Da aber auch bei prinzipiell gleichen Randbedingungen zu unterschiedlichen Zeiten aufgenommene Spektren sich geringfügig unterscheiden, haben wir zunächst die bereits vorhandene räumliche Struktur und die dazu passende sequenzielle Zuordnung optimal an die zu untersuchenden experimentellen Daten angepasst.

Dazu wurde mit der vorhandenen Struktur, dem vorhandenen spt- File und dem neuen experimentellen Spektrum für 298K ein erstes Mal automatisch zugeordnet. Mit dem Ergebnis dieser Zuordnung wurde einerseits eine benannte Masterliste erhalten, andererseits eine neue sequenzielle Zuordnung, die besser an das experimentelle Spektrum angepasst ist. Aus der benannten Masterliste (benanntes Spektrum) wurde ein Restraint-File für eine Strukturrechnung erzeugt, und damit eine neue 3D- Struktur für 298K (beste aus 1000 berechneten Strukturen). Diese neue Struktur zeichnete sich durch bessere Energiewerte aus (vgl. Abbildung 74), glich ansonsten nahezu der bereits vorhandenen, was auch erwartet war.

Diese neue Struktur für 298K nun verwendeten wir wieder dazu, das gleiche Spektrum noch einmal zuzuordnen. Daraus erhielten wir wieder eine sequenzielle Zuordnung und ein benanntes Spektrum. Diese Prozedur wurde insgesamt 4 Mal durchgeführt, mit dem Sinn, die im weiteren verwendete Ausgangsstruktur optimal an die herrschenden Gegebenheiten anzupassen und diese energetisch zu „beruhigen“.

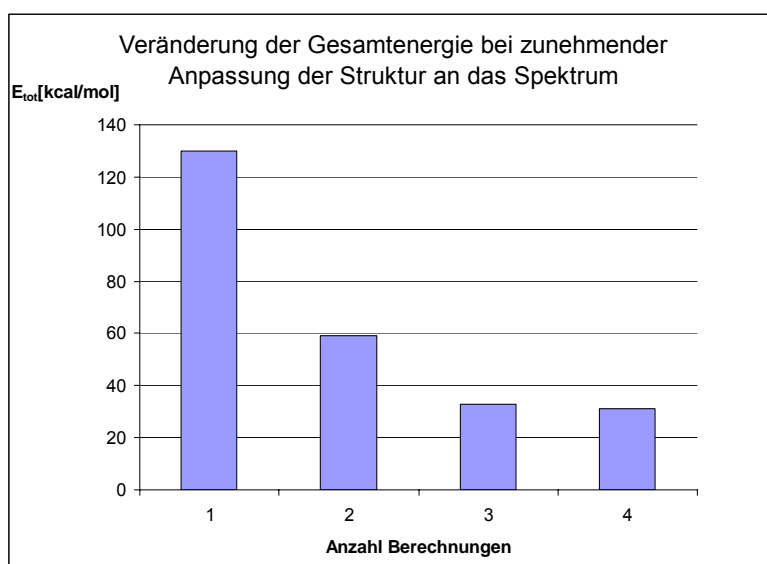


Abbildung 74: Energieminimierung der Ausgangsstruktur. Beim ersten Strukturrechnungsprozess beträgt die Pseudoenergie noch ca. 130 Kcal/mol, bei Schritt 3 und 4 pendelt sich diese auf ca. 30 Kcal/mol ein und verändert sich nicht weiter.



Wie Abbildung 75 zeigt, ändert sich die Struktur von Saratin nur geringfügig im Loop- Bereich während der erfolgten Anpassung an das experimentelle Spektrum. Die Energiewerte aber werden signifikant besser.

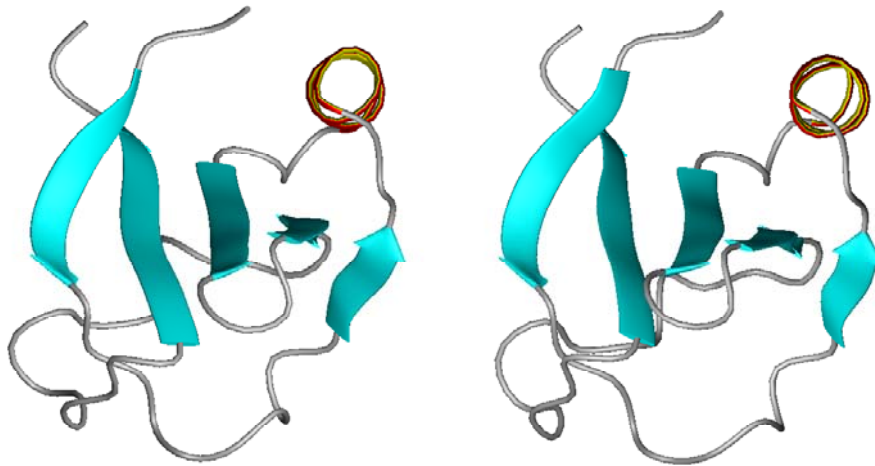


Abbildung 75: Räumliche Struktur von Saratin bei 298K. Links Struktur 298K\_1:  $E_{tot} = 130 \text{ kcal/mol}$ , rechts Struktur 298K\_4:  $E_{tot} = 31 \text{ kcal/mol}$ .

Nach erfolgter optimaler Anpassung der Struktur für 298K an das Experiment haben wir davon ausgehend in 1K- Schritten nach oben und unten zugeordnet. Das bedeutet, dass mit der Struktur und der sequenziellen Zuordnung von 298K jeweils das Spektrum für 297K und 299K zugeordnet wurden.

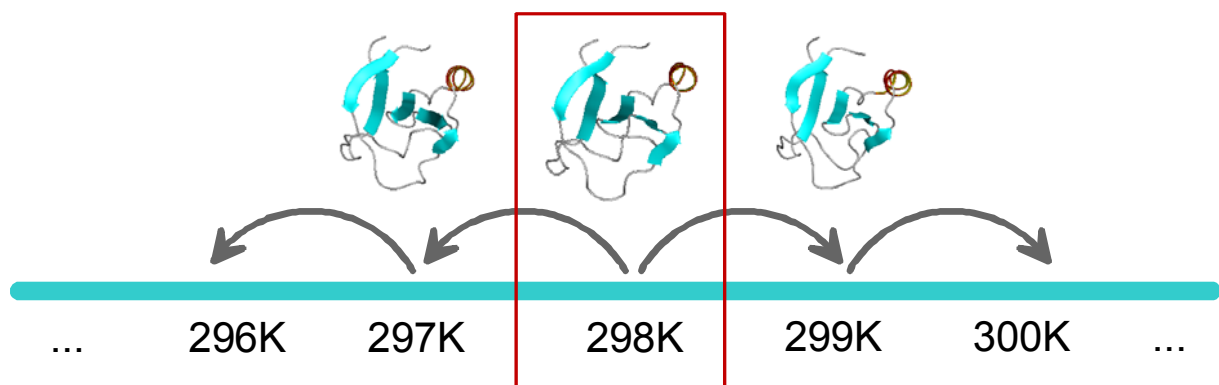
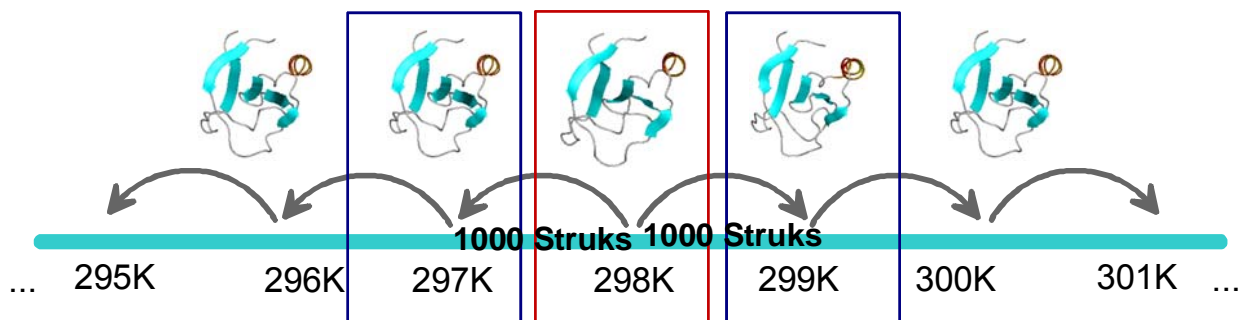


Abbildung 76: Ausgangspunkt der automatischen Zuordnung. Spektrum und Struktur für 298K. Damit werden schrittweise alle anderen Spektren zugeordnet.

Nach der automatischen Zuordnung wurden mittels der erhaltenen benannten Masterliste jeweils 1000 Strukturen berechnet. Die energetisch Besten aus dieser Strukturrechnung, sowie die neuen, durch den Programmablauf veränderten, sequenziellen Zuordnungen wurden dann für die weitere Zuordnung der Spektren für 296K und 300K herangezogen. So wurde



schrittweise den ganze gemessene Temperaturbereich zugeordnet, und für jede Temperatur eine Struktur erzeugt.



*Abbildung 77: Ausgehend von der Struktur und der Zuordnung von 298K werden jeweils 1000 Strukturen für die angrenzenden Temperaturen berechnet, mit den energetisch Besten daraus wird so weiterverfahren.*

Zur Berechnung der Strukturen kamen nur automatisch zugeordnete NOEs zur Verwendung, keine Zusatzinfos, wie z.B. H-Brücken oder Winkelbeschränkungen. Da die Anzahl der von AUNOAS benannten Signale pro Spektrum zwischen ca. 2000 und 3000 schwankte, waren auch mehr als genug Informationen für den Strukturrechenprozess vorhanden. (Zum Vergleich: manuell benannte 2D NOESYs bestehen aus ca. 1000 benannten Signalen).

Für die durchgeführten Strukturberechnungen wurden das standardmäßig eingestellte Annealing-Protokoll der CNS-Version 1.1, sowie die Defaultwerte für alle Kraftkonstanten verwendet. Bei der Berechnung der Strukturen mit CNS wurden auf die Distanzen zunächst 25% Fehler gegeben, mit einer unteren Grenze von 1,8 Angström, wie das auch bei manuell bestimmten Strukturen der Fall ist, um dem Strukturrechenprogramm zu ermöglichen, möglichst alle Randbedingungen zu erfüllen, und potentielle Fehler bei der Zuordnung auszugleichen. In einem weiteren Test wurden diese Fehlergrenzen auf 5% zurückgesetzt, um die Güte der automatischen Zuordnung zu überprüfen und auch, um die Beobachtungen über das Temperaturverhalten von Saratin zu festigen. Darauf wird später noch genauer eingegangen werden.

Nach der automatischen Zuordnung mit AUNOAS waren benannte Masterlisten (2D-Spektren) für alle untersuchten Temperaturen von 10°C bis 50°C vorhanden. Damit konnten wir für jede Temperatur eine räumliche Struktur von Saratin berechnen. Außerdem entstand für jede Temperatur eine an das jeweilige experimentelle Spektrum optimal angepasste sequenzielle Zuordnung. Diese Zuordnungen machten es zusätzlich möglich, das Verhalten der einzelnen chemischen Verschiebungen aller Atome abhängig von der Temperatur zu verfolgen.

Der Prozess von dem Beginn der Aufnahme der Daten, über die Prozessierung und Zuordnung, bis hin zur Berechnung der 41 Strukturen dauerte etwa 3 Monate, wovon auf das Aufnehmen der Spektren ca. 2 Wochen, auf das Prozessieren und Peakpicken 3 Wochen und das Berechnen der Strukturen ca. 4 Wochen (1000 Strukturen bei einer Rechenleistung

von 10 GHz dauerten ca. 9 Stunden) entfielen. Die eigentlich zeitraubende und langwierige Arbeit der Zuordnung der Spektren –inklusive der Vorbereitung der Rechenfiles für das Programm– belief sich auf durchschnittlich 1 Stunde pro Spektrum. Dies bedeutet eine enorme Erleichterung und Zeitersparnis, und ermöglicht auch neue, interessante, bisher wegen der immens langen Dauer jedoch unmögliche Anwendungen, wie z.B. Temperatur-Druck oder pH- Reihen.

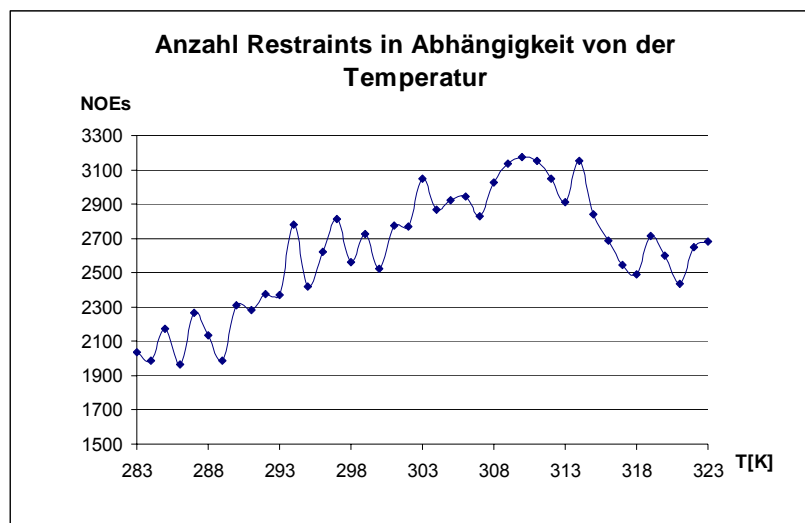
### 5.3. Strukturbestimmung

Aufgrund obiger Vorgehensweise erhielten wir zwei Mal 41 Saratin Strukturen. Die Strukturen wurden mit dem Programm CNS berechnet, es wurden dabei die automatisch zugeordneten NOE- Kontakte als Beschränkungen übergeben. Zusätzliche Angaben wurden dem Strukturrechenprogramm nicht gemacht. CNS simulierte ausgehend von einem ausgestreckten Strang pro Temperaturschritt 1000 Strukturen. Davon wurden die jeweils 10 energetisch Besten<sup>7</sup> ausgewählt und aus diesen mittels MOLMOL [Mol04] eine mittlere Struktur erzeugt.

Die für die Strukturrechnung verwendeten zugeordneten Signale schwanken abhängig von der Temperatur, der Geduld beim manuellen Peakpicken und auch dem automatischen Zuordnungsprozess zwischen 2000 und 3200 Signalen. Wie Abbildung 78 zeigt, spiegelt sich in der Anzahl der zugeordneten Signale deutlich wieder, dass bei kälteren Temperaturen weniger Signale zu beobachten sind. Bei sehr hohen Temperaturen haben wir in den Spektren teilweise schon Peakaufspaltungen durch J- Kopplung beobachtet, da die  $T_2$ -Relaxation aufgrund der geringer werdenden Korrelationszeit  $T_c$  vermindert war. Außerdem trat dort auch das vorhandene Rauschen stärker zu Tage. Der Bereich mit den meisten zugeordneten Signalen bewegt sich zwischen ca. 32°C und 42°C.

---

<sup>7</sup> Auswertung der im CNS- Logfile angegebenen Energieterme analog dem Programm *cnsstat* [Ste01].



*Abbildung 78: Anzahl der automatisch zugeordneten Signale abhängig von der Temperatur. Im Bereich bis zu 293K (20°C) sind deutlich weniger Signale zu beobachten, als bei höheren Temperaturen.*

Wie bereits in Kapitel 2.3 angesprochen, liefert die Berechnung der RMSD- Werte für die ausgewählten jeweils 10 Besten aus den berechneten 1000 Strukturen ein Maß für deren Unterschiedlichkeit. Da das Protein im Bereich der N- und C- Termini jedoch sehr stark beweglich ist [Dom04], haben wir zur Bestimmung des RMSD- Wertes einen Bereich von Aminosäure 4 bis Aminosäure 82 (von 103) betrachtet, in dem das Protein gut definiert ist. Der RMSD- Wert gibt einen Eindruck, wie gut eine Struktur durch die hineingesteckten Einschränkungen festgelegt wird, darf aber nicht überinterpretiert werden, da er lediglich ein Verständnis darüber vermittelt, wie stark die ausgewählten Strukturen „clustern“ nicht aber ein Maß dafür darstellt, wie „richtig“ sie sind. Die ermittelten Zahlenwerte hängen dabei empfindlich von den gewählten Potentialfunktionen, Kraftkonstanten und Fehlergrenzen für die Restraints ab, was Abbildung 79 und Abbildung 80 sehr anschaulich darstellen.

Bei weiten Fehlergrenzen von 25% bewegt sich der RMSD- Wert im Mittel über alle Temperaturen im Bereich um 0,8 Angström, bei engen Fehlergrenzen um 0,3 Angström. Zu beobachten ist, dass der RMSD- Wert bei den Strukturen der kälteren Temperaturen schlechter ist, als im wärmeren Temperaturbereich, was logisch ist, da dort weniger Signale vorhanden waren, und damit zur Strukturbestimmung auch weniger Signale herangezogen werden konnten, was zu einer relativ schlechteren Definition dieser Strukturen führt. Trotzdem sind alle beobachteten RMSD- Werte, insbesondere diejenigen der engen Fehlergrenzen, sehr gut, wenn man bedenkt, dass alle Strukturbestimmungen automatisch durchgeführt wurden.

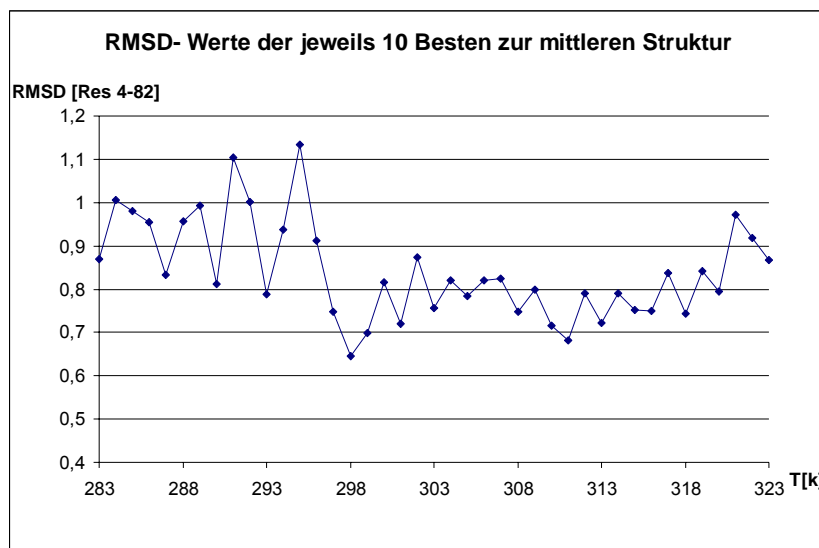


Abbildung 79: Darstellung der RMSD- Werte für die Saratinstrukturen, die mit weiten Fehlergrenzen von 25% berechnet wurden. Die Angabe des RMSD- Wertes erfolgt in Angström.

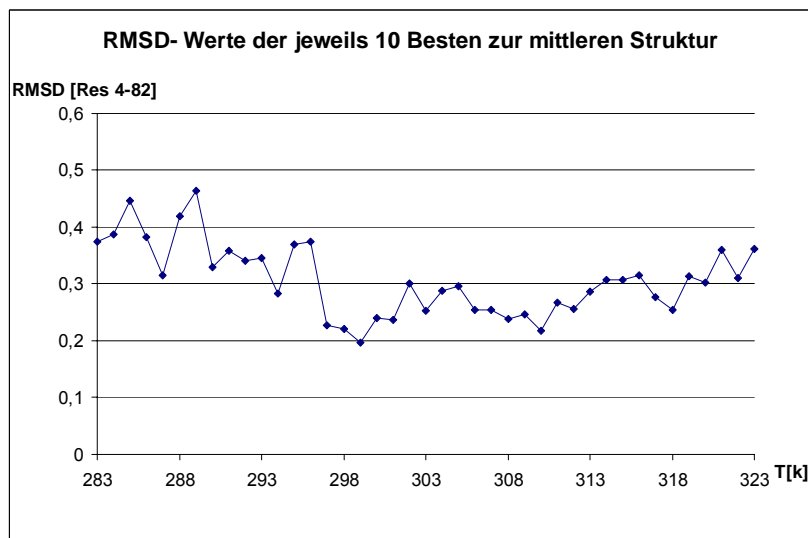


Abbildung 80: Darstellung der RMSD- Werte für die Saratinstrukturen, die mit engen Fehlergrenzen von 5% berechnet wurden. Die Angabe des RMSD- Wertes erfolgt in Angström.

Bezieht man nun auch noch die jeweiligen aus den Verletzungen der Einschränkungen resultierenden Gesamtenergien in die Betrachtung mit ein, so zeigt sich, dass die Strukturen ohne gravierende Verletzungen der NOE- Kontakte oder sonstiger Einschränkungen auch mit den sehr engen Fehlergrenzen von 5% berechnet werden können, was in Verbindung mit dem signifikant besseren RMSD- Wert und dem identischen Aussehen der Strukturen den

Schluss zulässt, dass einerseits bei der automatischen Zuordnung keine Fehler entstanden sind, und andererseits das Aussehen und die Lagebeziehungen der Strukturelemente der Realität entspricht.

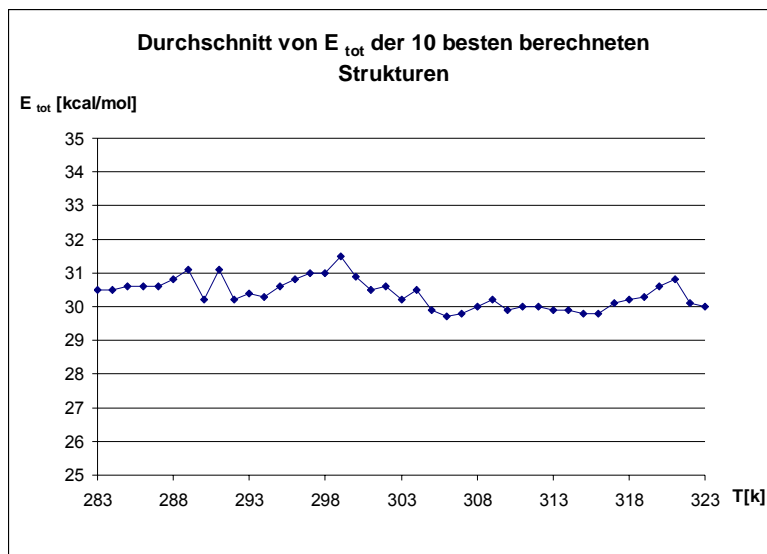


Abbildung 81: Darstellung der Gesamtenergien der mit weiten Grenzen berechneten Strukturen. Der Durchschnitt von  $E_{\text{tot}}$  wurde aus den Gesamtenergien der jeweils 10 besten aus 1000 Strukturen ermittelt.

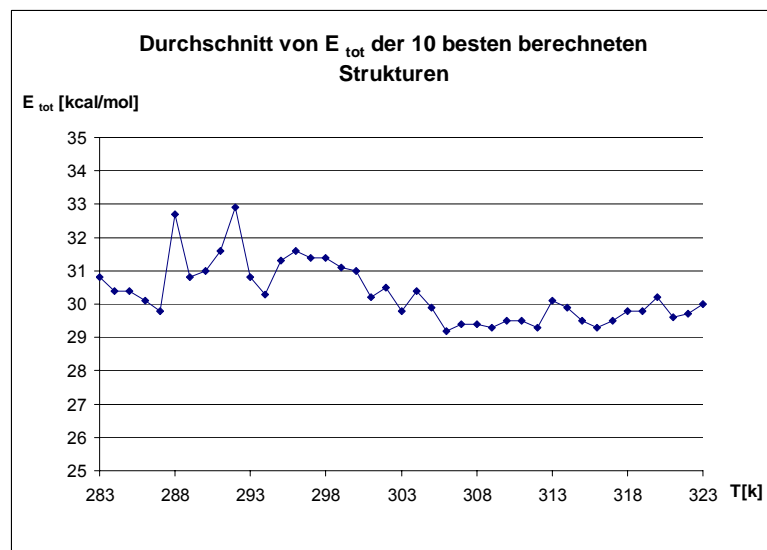


Abbildung 82: Darstellung der Gesamtenergien der mit engen Grenzen berechneten Strukturen. Der Durchschnitt von  $E_{\text{tot}}$  wurde aus den Gesamtenergien der jeweils 10 besten aus 1000 Strukturen ermittelt.

Die Gesamtenergiewerte schwanken in beiden Fällen um 30 kcal/mol, wobei der Beitrag der Verletzungsenergien aus NOEs generell in der Größenordnung um 0,01 kcal/mol liegt, was

bedeutet, dass die automatische Zuordnung keinerlei gravierende Fehler aufweist. Zudem sind auch die Gesamtenergien sehr gering und bestätigen eine korrekte Definition der Strukturen. Es kann sogar eine geringfügig geringere mittlere Gesamtenergie bei den Strukturen mit engen Fehlergrenzen beobachtet werden, was nur bei richtiger Zuordnung erwartet werden kann. Die geringfügig höheren Energiewerte bei kälteren Temperaturen rühren dabei aus einem geringeren  $T_c$ , welches eine erhöhte Spindiffusion bewirkt, welche wiederum zu ungenaueren Volumenwerten und damit zu höheren Gesamtenergien führen.

## 5.4. Ergebnisse

### 5.4.1. Räumliche Strukturen

Die Betrachtung der Strukturen zeigt, dass Saratin über den gesamten Temperaturbereich stabil ist (Hitzenaturierung erst ab mehrstündigem Erwärmen auf 75°C, Kältenaturierung nicht durchgeführt) und eine wohlgefaltete räumliche Struktur aufweist, die aus 5 antiparallelen  $\beta$ -Faltblättern und einer  $\alpha$ -Helix besteht. Es zeigt sich weiterhin, dass wie erwartet, die Loop-Bereiche beweglicher als die Sekundärstrukturelemente sind.

Was wir nicht beobachten konnten, war eine definierte Bewegungsrichtung einiger oder aller Teile der Struktur, stattdessen wurde eine gewisse Oszillation im Molekül sichtbar, die sich bei filmischer Betrachtung in einem „Pulsieren“ der 3D-Struktur bei zunehmender Temperatur äußert. Diese Tatsache konnte sowohl in den Strukturen, die mit weiten Fehlergrenzen berechnet wurden, als auch den Strukturen mit engen Fehlergrenzen beobachtet werden. Dies schließt aus, dass der beobachtete Effekt auf den Strukturberechnungsalgorithmus zurückzuführen ist (wie genau die Aminosäurekette räumlich gelegt werden darf, ist abhängig von den vorgegebenen Grenzen; je weiter diese sind, desto mehr Rauschen ist möglich), und lässt den Schluss zu, dass es sich um ein reell vorhandenes Phänomen handelt.

Abbildung 83 stellt alle erhaltenen mittleren Strukturen des untersuchten Temperaturbereiches als Bändermodell dar. Bei dieser Darstellung wird insbesondere die Faltung der Hauptkette und die Anordnung der Sekundärstrukturelemente betont. Um die Variationsbreite der Strukturen zu veranschaulichen, werden „Sausage“-Plots gewählt, bei denen die Hauptketten der gewählten Strukturen überlagert werden, und die Abweichung zum Mittelwert dargestellt wird (siehe unten). Eine Darstellung aller 41 Strukturen und der Vergleich von engen und weiten Fehlergrenzen bei der Strukturbestimmung findet sich im Anhang.

Es ist zu beobachten, dass sich die jeweils berechneten Strukturen kaum unterscheiden, was bei korrekter räumlicher Molekülstruktur auch nicht zu erwarten ist.

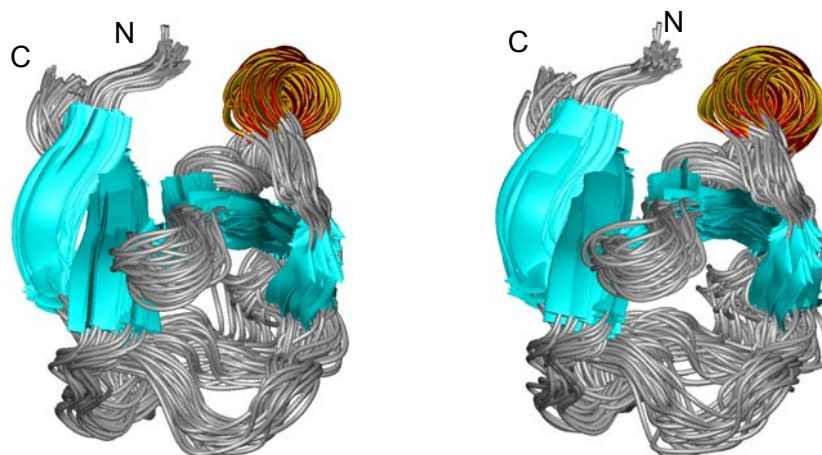


Abbildung 83: Alle berechneten mittleren Strukturen für den Temperaturbereich von 283K bis 323K in 1K-Schritten. Links: Strukturen berechnet mit weiten Fehlergrenzen (25%), rechts: Strukturen berechnet mit engen Fehlergrenzen (5%). Die mittleren Strukturen wurden jeweils aus einem Bündel der 10 Besten aus 1000 gerechneten Strukturen bestimmt.

Abbildung 84 zeigt die Sausage Plots von Saratin aus den mittleren Strukturen von 283K bis 323K in 1K-Schritten; die mittleren Strukturen wurden jeweils aus einem Bündel der 10 Besten aus 1000 gerechneten Strukturen bestimmt. Der Plot stellt dar, in welchem Bereich sich das Molekül bei zunehmender Erwärmung bewegt. Wie man deutlich sieht, sind die Loop- Bereiche beweglicher, wogegen z.B. besonders die innen liegenden Faltblätter sehr ortstabil sind. Dennoch ist das Molekül über den ganzen Bereich gut definiert und behält seine räumliche Struktur relativ unverändert bei. Der RMSD- Wert über alle 41 Strukturen sowohl mit engen, als auch mit weiten Grenzen beläuft sich auf ca. 1,4 Angström, deutlich mehr, als die RMSD- Werte der einzelnen mittleren Strukturen, was zeigt, dass eine Bewegung, wenngleich auch nicht gerichtet, stattfindet.

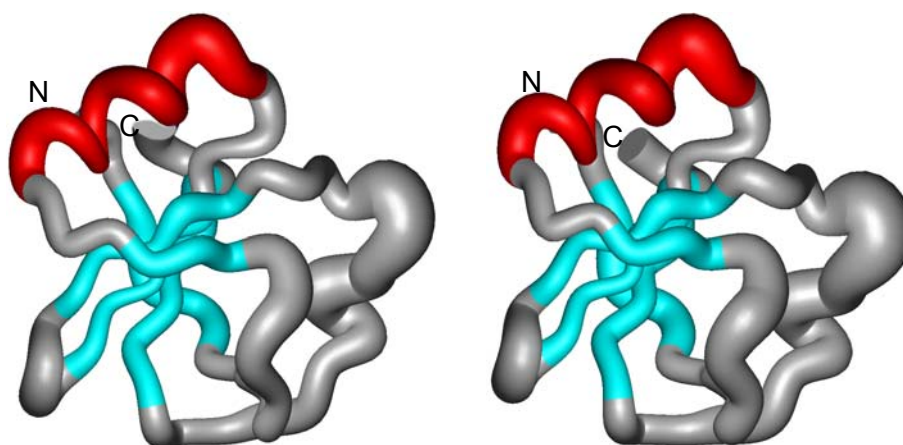
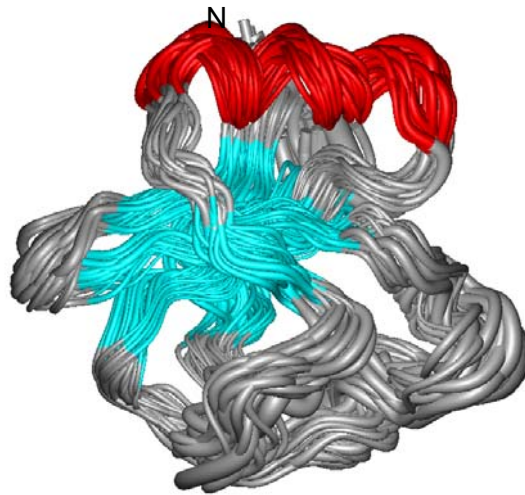


Abbildung 84: Sausage Plots über alle berechneten Strukturen. Links: Strukturen berechnet mit weiten Fehlergrenzen (25%), rechts: Strukturen berechnet mit engen Fehlergrenzen (5%).



Betrachtet man die einzelnen Sausage- Plots aller Temperaturen, so sieht man, dass diese im Vergleich zur gesamten Änderung sehr scharf definiert sind.

Das induziert für beobachtete Effekte, dass diese nicht durch Rauschen oder große Variationsbreite der einzelnen Strukturen begründet sind.



*Abbildung 85: Sausage Plots der 10 besten aus 1000 für jede Temperatur berechneten Strukturen. Man sieht die gute Definition der einzelnen mittleren Strukturen (insbesondere im Bereich der Sekundärstrukturelemente) und den dazu vergleichsweise großen Bereich der Änderungen während der Veränderung der Temperatur.*



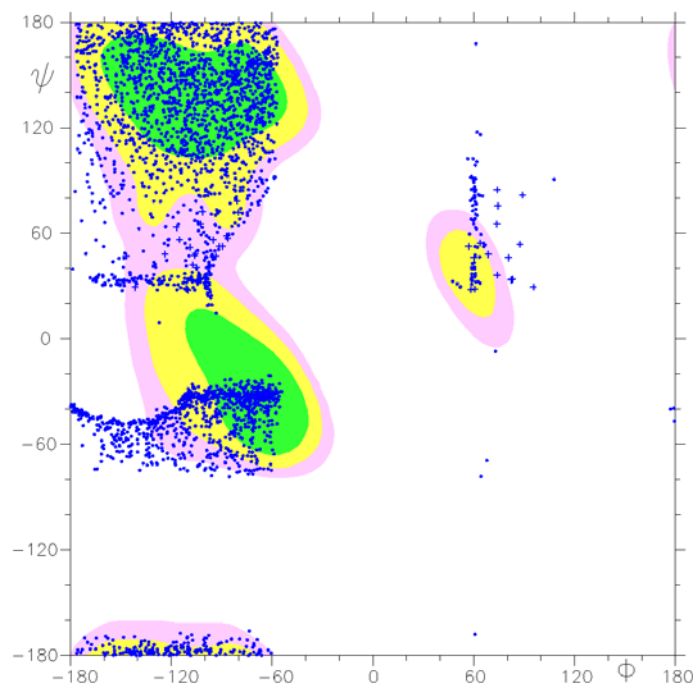


Abbildung 86: Lage der Diederwinkel  $\Phi$  und  $\Psi$  im Ramachandran-Plot aller 41 mittleren Strukturen. Glycine sind durch ein Kreuz gekennzeichnet. Deutlich erkennbar ist eine Häufung der Werte in den Winkelbereichen, die einer definierten Sekundärstruktur entsprechen. Für die Glycin- Reste ist ein erweiterter Winkelbereich zugänglich.

Abbildung 86 stellt die Hauptketten- Torsionswinkel der mittleren Struktur jedes Temperaturschrittes in einem Ramachandran- Diagramm dar. Mit Ausnahme der Aminosäuren aus den restlichen ungefalteten Bereichen des N- und C- Terminus, sowie der Loops und dem Ende der Helix (letzte Aminosäure), das, wie in den Strukturen beobachtet, nicht sehr stabil ist, befinden sich die Winkelpaare fast ausschließlich in den sterisch günstigen Regionen, welche für  $\alpha$ - Helices und  $\beta$ - Faltblätter charakteristisch sind.

Betrachtet man die erzeugten Strukturen im Ablauf der Temperaturänderung, so beobachtet man, wie bereits oben angeschnitten, einen Effekt, der ein „Schwingen“ der Struktur verursacht. Es ist nun aufschlussreich diese Beobachtung weiter zu untersuchen, um auszuschließen, dass es sich um Rauschen oder Artefakte, die aus dem automatischen Zuordnungsprozess stammen, handelt.

Unser erster Test bestand darin, alle Strukturen wie schon ausgeführt, mit sehr engen Fehlergrenzen zu berechnen. Des weiteren haben wir in einem nächsten Versuch die CA-Atome der Hauptkette untersucht. Zuerst wurden die Abstände aller sequenzieller CA-Atome bei allen Temperaturen berechnet, graphisch aufgetragen und ausgewertet.

Ziel dieser Arbeit war es, zu erkennen, ob die beobachtete Oszillation einer bestimmten Periodizität folgt, oder ob sich statistisch mit jeder Änderung der Temperatur auch die

Abstände der einzelnen Atome untereinander ändern. Untersucht wurden wiederum die jeweiligen mittleren Strukturen.

Abbildung 87 und Abbildung 88 zeigen zwei repräsentative Beispiele der Untersuchungen.

#### CA 8/ CA9

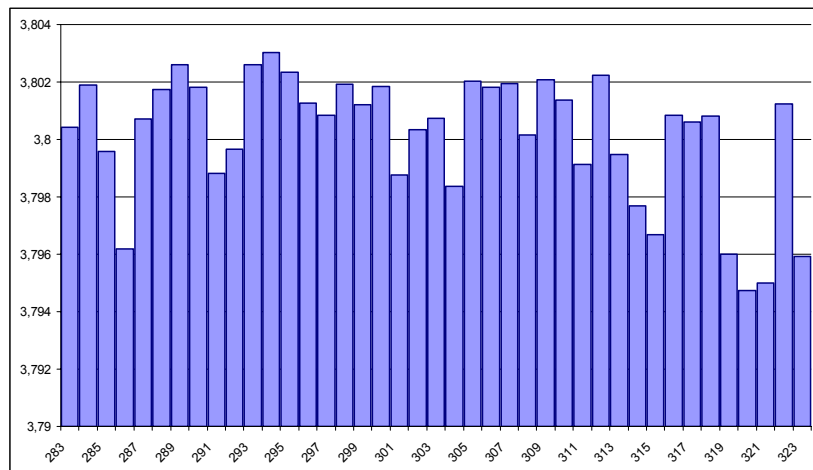


Abbildung 87: Darstellung der Veränderung der Abstände der CA- Atome der Aminosäuren 8 und 9. Auf der x- Achse ist die Temperatur in K, auf der y- Achse die Distanz in Angström aufgetragen.

#### CA 17/ CA 18

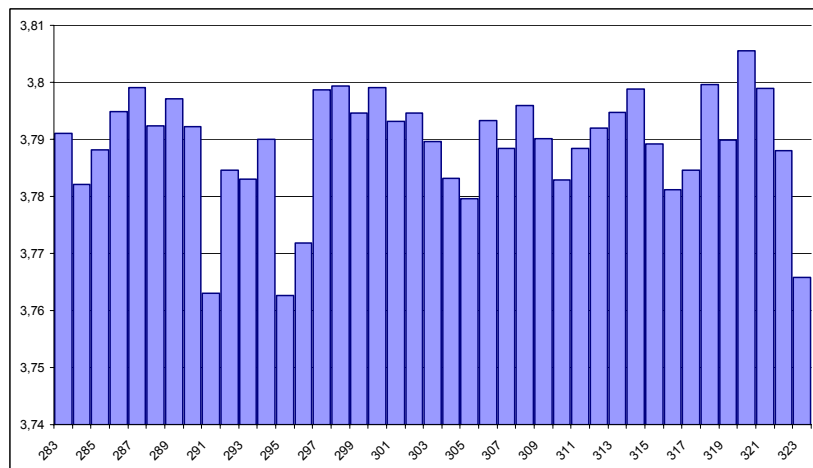


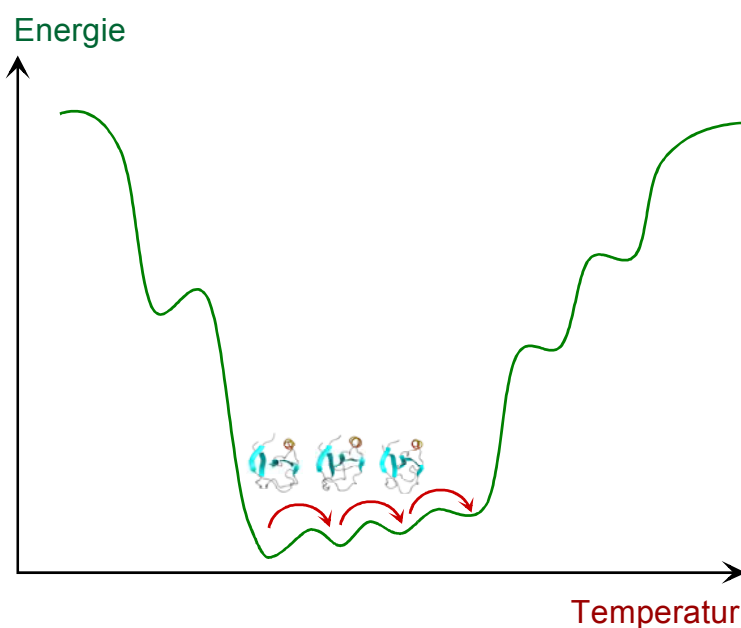
Abbildung 88: Darstellung der Veränderung der Abstände der CA- Atome der Aminosäuren 17 und 18. Auf der x- Achse ist die Temperatur in K, auf der y- Achse die Distanz in Angström aufgetragen.

Wie man sieht, weisen die Änderungen in den Abständen eine gewisse Regelmäßigkeit auf, und scheinen sich nicht statistisch zu verhalten, oder mit jeder Temperaturänderung willkürlich zu verändern.

Es ist alle 3- 5 Temperaturschritte eine Erhöhung, bzw. Verringerung der betrachteten Abstände sichtbar, was auch auf eine definierte Periodizität schließen lässt.

Die Vorstellung dazu, die auch bereits allgemein diskutiert wird [Kal04] ist, dass auch bei wohlgefalteten Molekülen (am Boden des „Faltungstrichters“ siehe Abbildung 89) unterschiedliche Energieniveaus (Feinstruktur) existieren, (der „Trichter“ also nicht glatt, sondern uneben ist) in die die Moleküle bei Änderung der Temperatur teilweise bis vollständig überführt werden.

Was man letztlich bei einer bestimmten Temperatur beobachtet, sind Mittelwerte aus den unterschiedlichen Konformationen, die die vorhandenen Moleküle überwiegend einnehmen.



*Abbildung 89: Vorstellung zur Faltung eines Proteins. Bei Beginn der Faltung strebt das Molekül zum Zustand der geringsten Energie. Aber auch am „Grund“ dieses „Faltungstrichters“ existieren noch geringe Energieunterschiede, die bei Erhöhung der Temperatur überwunden werden und wobei die Population der Moleküle zwar wohlgefaltete aber andere Konformationen einnimmt.*

Natürlich sind die untersuchten Daten nicht frei von Rauschen, was bedeutet, dass die Beobachtungen nicht beliebig genau durchgeführt werden können, genauso wenig, wie eine Aussage darüber möglich ist, wie die „genauen“ räumlichen Strukturen bei einer bestimmten Temperatur aussehen, da man in der NMR immer Mittelwerte beobachtet. Da aber die Oszillation deutlich über 1° Schritte hinausgeht, kann man in diesem Fall davon ausgehen, dass ein anderer Effekt daran beteiligt ist.

### **Ausblick, weitere Arbeiten**

Um die dargestellten Vermutungen weiter zu sichern und zu untersuchen, haben wir auch CA- Atome weiter voneinander entfernter Aminosäuren betrachtet. Bei diesen dürfte die beobachtete Oszillation noch stärker zutage treten, als bei benachbarten Atomen. Erste

Versuche zeigen, dass diese Annahme korrekt ist, und gerade diese Atome eine relativ deutliche Periodizität aufweisen.

Abbildung 90 und Abbildung 91 zeigen zwei Beispiele dieser Untersuchungen. Vor allem in Abbildung 90 ist eine deutliche Schwingung in den Abständen zu erkennen.

### **CA(i)/ CA(i+4)**

CA 36/ CA 40

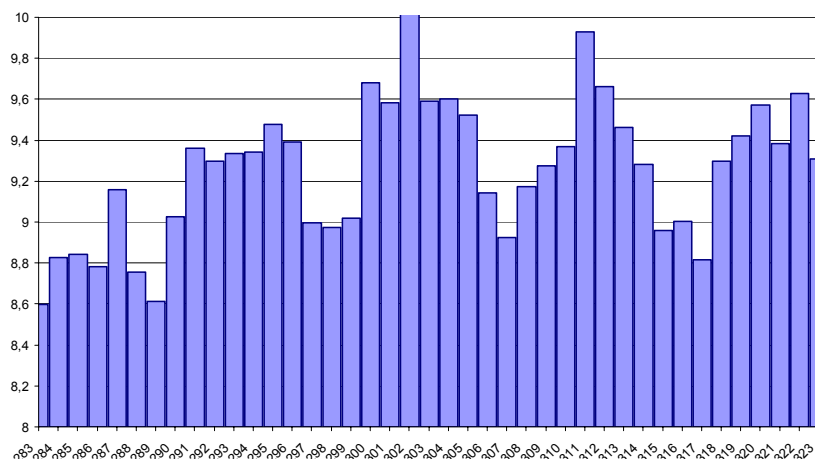


Abbildung 90: Darstellung der Veränderung der Abstände der CA- Atome der Aminosäuren 36 und 40. Auf der x- Achse ist die Temperatur in K, auf der y- Achse die Distanz in Angström aufgetragen. Deutlich zu erkennen sind Schwingungen, die etwa 9 Temperaturschritte umfassen.

### **CA(i)/ CA(i+20)**

CA 22/ CA 42

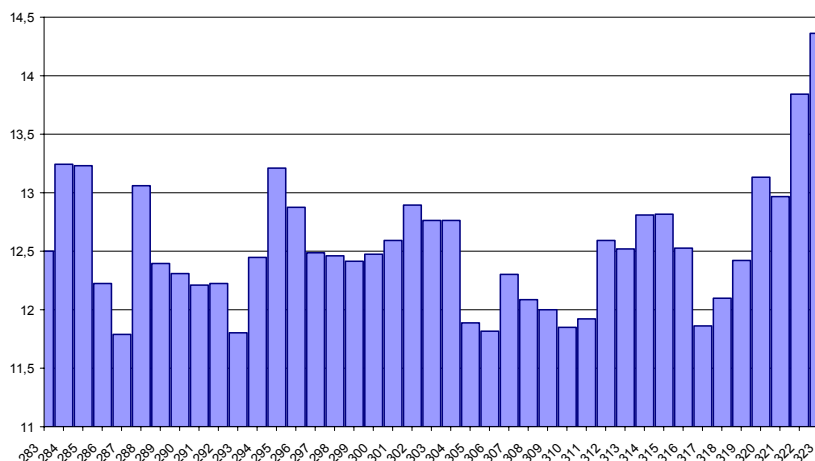


Abbildung 91: Darstellung der Veränderung der Abstände der CA- Atome der Aminosäuren 22 und 42. Auf der x- Achse ist die Temperatur in K, auf der y- Achse die Distanz in Angström aufgetragen. Auch hier sind Schwingungen zu erkennen die ca. 7 Temperaturschritte umfassen.

In einem weiteren Schritt sollen einerseits diese Untersuchungen auf alle Atome ausgedehnt werden, wobei unterschiedliche sequenzielle Abstände der Atome zu wählen sind, um die Oszillationen zu untersuchen. Andererseits sollte darüber hinaus im Anschluss eine mathematische Beschreibung der beobachteten Periodizität gefunden werden.

Dies dürfte zu einer Sicherung und einem besseren Verständnis der Strukturschwingungen führen.

#### 5.4.2. $H^N$ - Shifts

Aus der Arbeitsweise von AUNOAS und den damit durchgeführten Zuordnungen erhielten wir auch 41 Listen von sequenziellen Zuordnungen für die experimentellen Spektren bei den gegebenen Temperaturen. Dadurch wurde es uns möglich, eine Betrachtung aller  $H^N$ - Shifts über den Temperaturbereich von 283K bis 323K in 1K- Schritten durchzuführen, und die Änderungen der chemischen Verschiebungen bei Änderung der Temperatur zu verfolgen.

Die chemische Verschiebung von Amidprotonen wird stark von Wasserstoffbrückenbindungen zu anderen Protonen beeinflusst, und ist abhängig von deren relativen Abstand und der Orientierung [Wag83, Asa95]. Man nimmt an, dass ein zunehmender Abstand von H und O, oder eine Abweichung der Bindung von der Planarität in einer Veränderung der chemischen Verschiebung zu kleineren ppm- Werten resultiert. Eine Erhöhung der Temperatur bedeutet eine erhöhte Beweglichkeit der an einer Wasserstoffbrückenbindung beteiligten Atome, und dadurch wird im allgemeinen auch der Abstand beider Partner vergrößert, was eine Veränderung der chemischen Verschiebung zur Folge hat. Die Veränderung ist gering, wenn die beteiligte H-O- Bindung stark ist, wodurch der Grad der Veränderung der chemischen Verschiebung mit der Temperatur ein Maß dafür darstellt, wie lang und stark eine intramolekulare Wasserstoffbrückenbindung ist [And97, Bax97, Cier01].

Proteine dehnen sich nach gängiger wissenschaftlicher Meinung linear mit der Temperatur aus [Fra87, Til92]. Diese lineare Ausdehnung führt zu Wasserstoffbrückenbindungen, deren Länge ebenso proportional zur Temperatur zunimmt, und dabei bei starken H-O- Bindungen eine Änderung der chemischen Verschiebung um ca. 0,001ppm pro K hervorruft [Cor02]. Wie schon öfter ausgeführt wurde [And97, Bax97], ist die Änderung der chemischen Verschiebung bei vielen Amidprotonen in guter Näherung linear. Trotzdem gibt es einige Fälle, in denen nachgewiesen wurde, dass das Temperaturverhalten der Shifts eine Krümmung aufweist [Bax98, Wil03].

Unsere Hoffnung bei einer genügend großen Anzahl von ausreichend guten Datenpunkten ist es, eine Vorhersage mittels eines mathematischen Modells über das weitere Verhalten der chemischen Verschiebungen bei anderen Temperaturen treffen zu können. Dieses könnte für zukünftige Strukturbestimmungsbemühungen außerordentlich hilfreich sein.

Dazu haben wir zunächst alle  $H^N$ - Shifts des Proteins gesammelt und aufgetragen. Als erstes haben wir die linearen und quadratischen Koeffizienten des Temperaturverhaltens ermittelt. Es zeigte sich, dass die chemischen Verschiebungen mit zunehmender Temperatur generell

einem kleineren ppm- Wert zustreben<sup>8</sup> und auch in erster Näherung linear sind. Dies steht in Einklang mit bereits veröffentlichten Arbeiten und gängigen wissenschaftlichen Ansichten. [Bax97, Wil92, Ref03] Gleichzeitig wiesen jedoch leichte Abweichungen von exakter Linearität, die durch den weiten untersuchten Bereich und die Güte der Daten ermittelt werden konnten, darauf hin, dass ein anderes, als das lineare Modell zur Beschreibung des Temperaturverhaltens herangezogen werden muss.

#### 5.4.2.1. Lineare und quadratische Koeffizienten

Die Auftragung der linearen Koeffizienten zeigt in Loops generell einen größeren Wert, was mit der stärkeren Beweglichkeit in diesen Bereichen korreliert ist. Grundsätzlich sind außerdem alle Koeffizienten, bis auf die der  $H^N$ - Atome der Aminosäuren 49 und 50 negativ. Es zeigt sich weiterhin, dass in den besonders stabilen Bereichen im Inneren der Faltblätter (vgl. Abbildung 84) die kleinsten Änderungen der chemischen Verschiebungen auftreten. Abbildung 92 zeigt die Auftragung der linearen Koeffizienten in Abhängigkeit von der Sekundärstruktur. Dabei sind Loop- Bereiche grau,  $\beta$ - Faltblätter hellblau und  $\alpha$ - Helix- Bereiche rot eingefärbt.

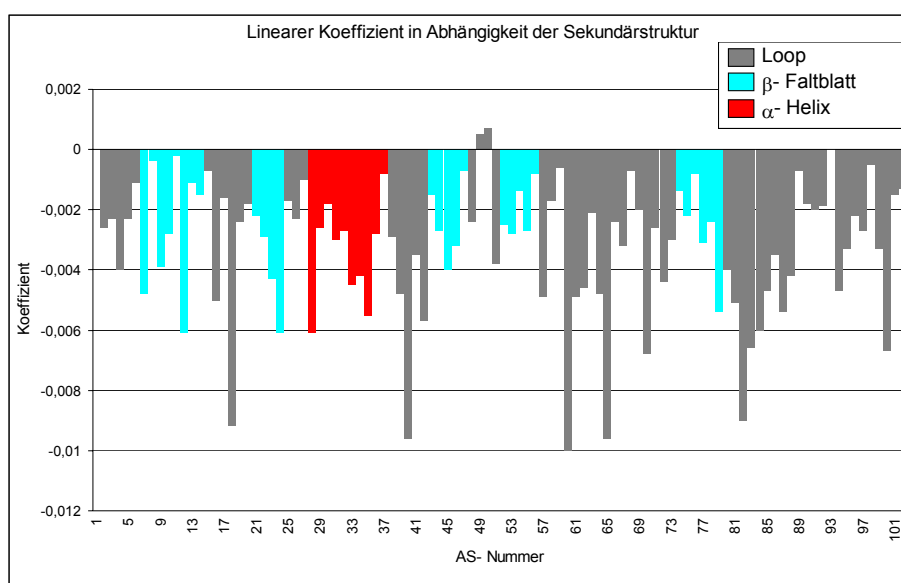
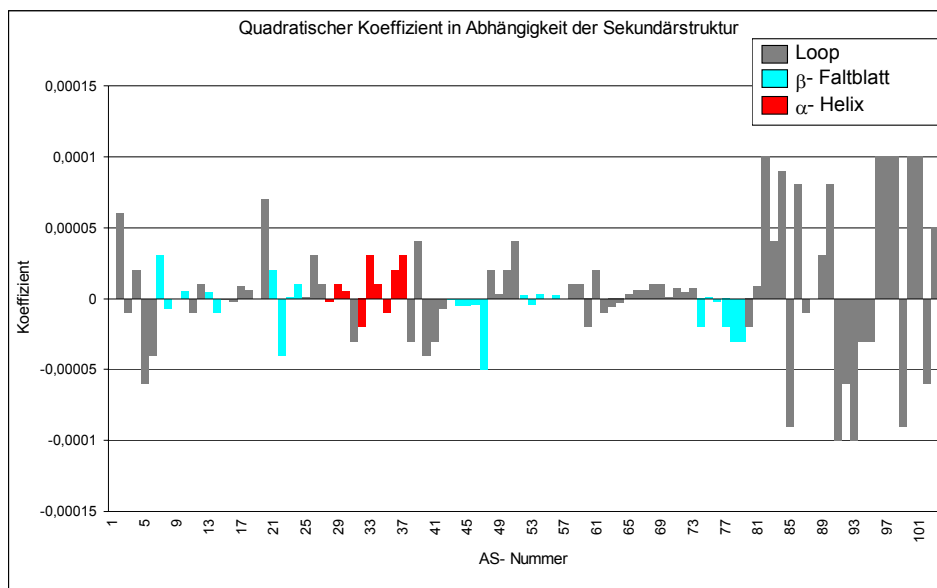


Abbildung 92: Auftragung der linearen Koeffizienten in Abhängigkeit von der Sekundärstruktur. Loop- Bereiche sind grau,  $\beta$ - Faltblätter hellblau und  $\alpha$ - Helix- Bereiche rot gekennzeichnet. Man sieht generell größere lineare Koeffizienten in den beweglichen Loop- Bereichen.

<sup>8</sup> Eine Ausnahme dabei bilden die  $H^N$ - Shifts der Aminosäuren 49 und 50, darauf wird weiter unten eingegangen.

Betrachtet man dagegen die Auftragung der quadratischen Koeffizienten, so zeigt sich keinerlei Korrelation mit Sekundärstrukturelementen, auch sind die Koeffizienten mit vier Nachkommastellen außerordentlich gering, was die Beobachtung widerspiegelt, dass das Verhalten der chemischen Verschiebungen, wie schon oben erwähnt, nahezu linear ist.

Die größten quadratischen Koeffizienten sind am C- Terminus zu beobachten, der allerdings äußerst beweglich ist, und, wie die Strukturrechnungen gezeigt haben, keine definierte Struktur aufweist. Dort sind auch die Daten teils verrauscht, was teilweise zu den beobachteten höheren quadratischen Koeffizienten führt.



*Abbildung 93: Auftragung der quadratischen Koeffizienten in Abhängigkeit von der Sekundärstruktur. Loop- Bereiche sind grau,  $\beta$ - Faltblätter hellblau und  $\alpha$ - Helix- Bereiche rot gekennzeichnet. Man sieht generell keine Korrelation mit Sekundärstrukturelementen*

Einer genaueren Betrachtung wurden die Aminosäuren 49 und 50 unterzogen, die als einzige einen positiven linearen Koeffizienten aufweisen. Allerdings ist dieser mit 0,005 und 0,007 relativ klein (Darstellung des Verlaufs der Shiftänderung im Anhang).

Aminosäure 49 (VAL) und 50 (ASN) befinden sich im Turn der  $\beta$ - Faltblätter 3 und 4 und weisen eine relativ exponierte Lage auf. Abbildung 94 veranschaulicht die Lage der Aminosäuren im Bändermodell, im raumfüllenden Kalottenmodell und in einer Oberflächendarstellung.

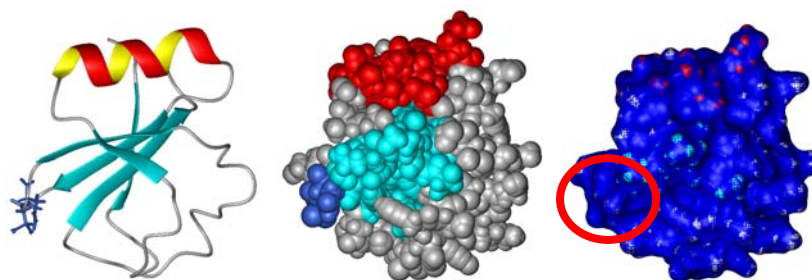


Abbildung 94: Darstellung von Saratin als Ribbon- Modell (links), als Kalottenmodell (mitte) und in Oberflächendarstellung (rechts). Die einzigen Aminosäuren mit positiven linearen Koeffizienten, Valin49 und Asparagin50, sind blau hervorgehoben, bzw. rot eingekreist.

In allen Darstellungen ist dieser Bereich exponiert und leicht zugänglich. Der erste Verdacht, es könnte sich ein aromatischer Ring in der Nähe befinden, der das Verhalten beeinflusst und evtl. strukturelle Änderungen hervorruft, bewahrheitete sich nicht. In der Sequenz und auch räumlich benachbart ist lediglich ein positiv geladener Lysin- Rest (ASR 47- THR 48- VAL 49- ASN 50- LYS+ 51- GLU 52).

#### 5.4.2.2. Darstellung und Beschreibung des Verhaltens der $H^N$ - Shifts

Wie bereits eingangs erwähnt, zeigten die beobachteten  $H^N$  - Shifts ein nahezu, aber nicht einwandfreies lineares Verhalten. Anhand der folgenden Abbildungen soll dieses exemplarisch illustriert werden. Für die Abbildungen gilt: Blau: Faltblatt, rot: Helix, farblos: Loop. Der Großteil aller anderen untersuchten chemischen Verschiebungen findet sich im Anhang.

$H^N$  31

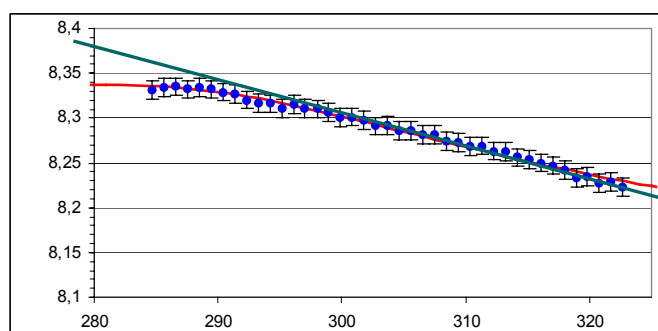


Abbildung 95: Chemische Verschiebung bei  $H^N$  31. Die in dunkelgrün angenäherte Gerade illustriert das nicht exakt lineare Verhalten der Datenpunkte. Auf der x- Achse ist die Temperatur in K, auf der y- Achse die Verschiebung in ppm aufgetragen. Die Fehlerbalken repräsentieren die digitale Auflösung der experimentellen Spektren, die rote Kurve stellt den Fit der Daten nach einem Zwei- Zustands- Modell (vgl. folgende Seiten) dar.



**H<sup>N</sup> 46**

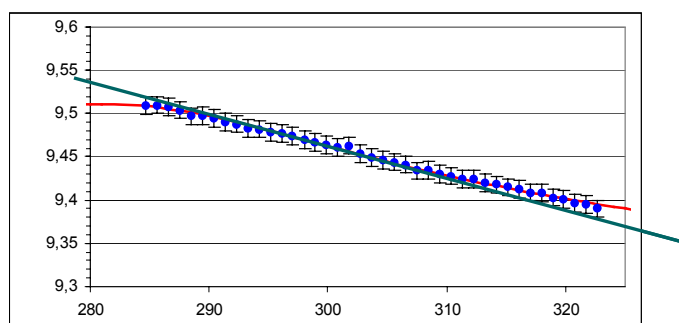


Abbildung 96: Chemische Verschiebung bei H<sup>N</sup> 46. Analog Abbildung 95.

**H<sup>N</sup> 79**

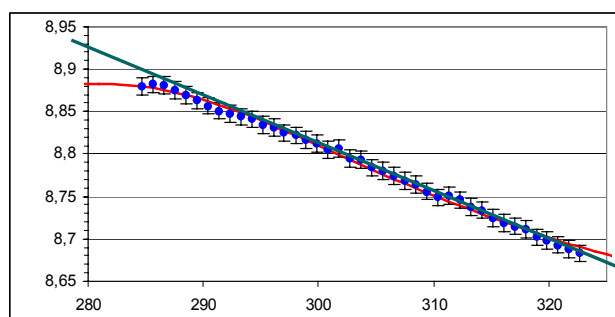


Abbildung 97: Chemische Verschiebung bei H<sup>N</sup> 79. Analog Abbildung 95.

**H<sup>N</sup> 83**

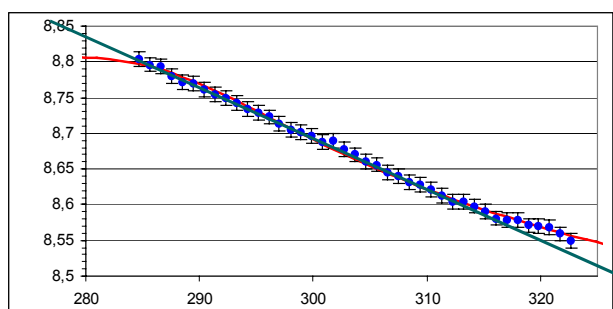


Abbildung 98: Chemische Verschiebung bei H<sup>N</sup> 83. Analog Abbildung 95.

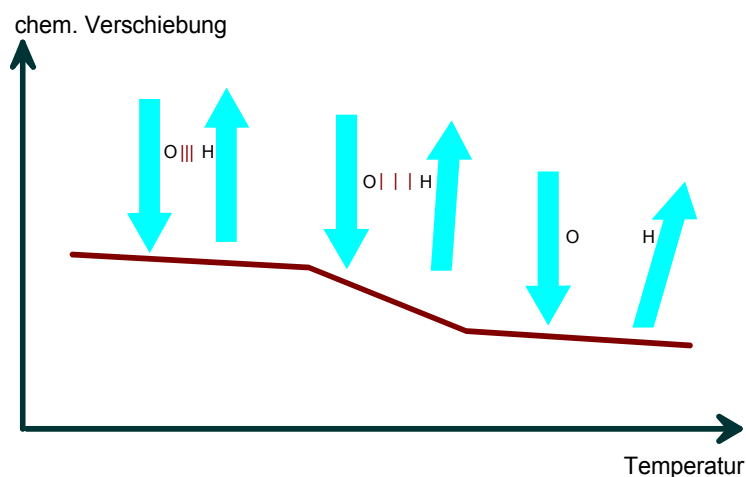
Alle Kurven zeigen eine leichte Abweichung von der Linearität und die hohe Qualität der Daten, mittels derer es möglich ist, detaillierte Untersuchungen anzustreben. Der in den Abbildungen bereits integrierte Fit beschreibt die Daten gut und mit geringer Abweichung.

### **Idee des Zwei- Zustands- Modell**

Dem beobachteten Verhalten der Shifts haben wir ein Zwei- Zustands- Modell zugrunde gelegt, das auf der Annahme beruht, dass eine Wasserstoffbrückenbindung bei

zunehmender Temperaturerhöhung und damit verbundener struktureller Änderung eine „Dehnung“ erfährt, die mit Shiftänderungen verbunden ist. Beim Aufbrechen der H-O-Bindung ergeben sich stärkere Änderungen der chemischen Verschiebung. Die Endpunkte der in diesem Modell angenommenen Verlaufskurve der Shifts, liegen auf der einen Seite bei „H-O- Bindung komplett geschlossen“ und auf der anderen Seite bei „H-O- Bindung nicht mehr vorhanden“. Die realen Endpunkte befinden sich sicherlich an den Punkten der Kälte- und Hitzedenaturation. Abbildung 99 veranschaulicht das Modell.

In diesem Zusammenhang wurden bereits von Williamson [Wil03] Untersuchungen angestellt, die einen gekrümmten, konvexen, bzw. konkaven Verlauf der chemischen Verschiebung vorhersagen, wenn die Daten nach einem linearen Modell gefittet werden, und dann dieser lineare Anteil von den vorhandenen Daten abgezogen wird. Dabei wurde von ihm ein Unterschied von 0,5ppm von Grundzustand und alternativem Zustand und eine Änderung der freien Energie bei einem Bruch der H-O- Bindung von 1kcal/mol angenommen. Williamson führte die Untersuchungen bei sieben unterschiedlichen Temperaturen, in einem Bereich von 280K bis 352K unter Anwesenheit unterschiedlicher Konzentrationen eines denaturierenden Agens (Guanidinhydrochlorid) durch.



*Abbildung 99: Idee des Zwei- Zustands- Modells. Ist die Wasserstoffbrückenbindung geschlossen, ergeben sich bei zunehmender Erwärmung nur leichte Änderungen der chemischen Verschiebung, die in einer „Dehnung“ begründet sind. Entfernen sich die beteiligten Atome zu weit voneinander bricht die Bindung auf, was in einer verhältnismäßig starken Shiftänderung resultiert. Ist schließlich keine H- O- Bindung mehr vorhanden, sind die weiteren Auswirkungen auf die chemischen Verschiebungen bei zunehmender Temperatur wieder geringer.*

Aufgrund dieser Annahmen wurde begonnen, die Daten mittels einer leicht veränderten Formel für ein chemisches Zwei- Zustands- Modell zu fitten [Wil03]. Im Gegensatz zu den Arbeiten Williamsons wurden hier jedoch alle Parameter frei gelassen, da die Hoffnung bestand, die Daten aufgrund ihrer Vielzahl und Qualität auch so ausreichend gut annähern

zu können. Außerdem wurde auf die Differenzbildung mit dem linearen Modell verzichtet, da auch an den reinen Daten die Abweichung von der Linearität gut zu erkennen ist.

$$\delta(T) = (\delta_A - \delta_B) \frac{e^{-\frac{\Delta G_0}{R(T-T_0)}}}{1 + e^{-\frac{\Delta G_0}{R(T-T_0)}}} + \delta_B$$

*Gleichung 9: Verwendete Funktion zur Beschreibung des Verhaltens der  $H^N$ -Shifts.*

wobei:

- $\delta(T)$ : beobachtete chemische Verschiebung bei einer bestimmten Temperatur [ppm]  
 $(\delta_A - \delta_B)$ : Unterschied der chemischen Verschiebungen am Anfangs- (H-O- Brücke geschlossen) und Endpunkt (H-O- Brücke geöffnet) ( $\Delta\delta$ ) [ppm]  
 $R$ : allgemeine Gaskonstante [J/mol K]  
 $\Delta G_0$ : Änderung der freien Enthalpie [J]  
 $T$ : aktuelle Temperatur [K]  
 $T_0$ : Temperatur am „Nullpunkt“ [K]  
 $\delta_B$ : Shift [ppm] am Endpunkt (H-O- Brücke vollständig geöffnet)

Eine Schwierigkeit in dieser verwendeten Formel stellt die Definition des Parameters  $T_0$  dar. Es ist nicht ohne weiteres erkenn- und begründbar, welchen Wert dieser annehmen sollte. Einerseits könnte er den absoluten Nullpunkt bei 0K repräsentieren, andererseits den Gefrierpunkt bei 0°C, also 273K. Auch denkbar wäre es, dass in diesem Zusammenhang der Nullpunkt bezüglich der Temperatur den Punkt der Kältedenaturation des verwendeten Proteins angibt. Um den Einfluss zu untersuchen, den die Wahl von  $T_0$  auf die Fitfunktion ausübt, wurden die Daten zunächst mit unterschiedlichen Werten von  $T_0$  angenähert<sup>9</sup>. Dabei haben wir untersucht, welche Kombinationen aus  $\Delta G_0$  und  $\Delta\delta$  ( $\delta$ Shift) bei einem gegebenen  $T_0$  die besten Werte in der Summe der quadratischen Abweichungen aufweisen, was ein Maß dafür ist, wie gut die Daten durch den verwendeten Fit erklärt werden.

### ***Einfluss von $T_0$ auf den Fit der $H^N$ - Daten, beispielhaft dargestellt am Atom $H^N$ 31***

Die folgenden Abbildungen illustrieren das Ergebnis beispielhaft an einigen Temperaturen für die chemischen Verschiebungen des Atoms  $H^N$  31. Weitere Temperaturschritte für  $T_0$  finden sich im Anhang dargestellt

<sup>9</sup> Anmerkung: Aufgrund der vielen freien Fit- Parameter in dieser Gleichung war es nicht möglich, auch noch  $T_0$  innerhalb eines Rechengangs annähern zu lassen.

Quadratische Abweichung des Fits bei unterschiedlichen Kombinationen von  $\Delta\delta$  und  $\delta G$  bei  $T_0=0K$

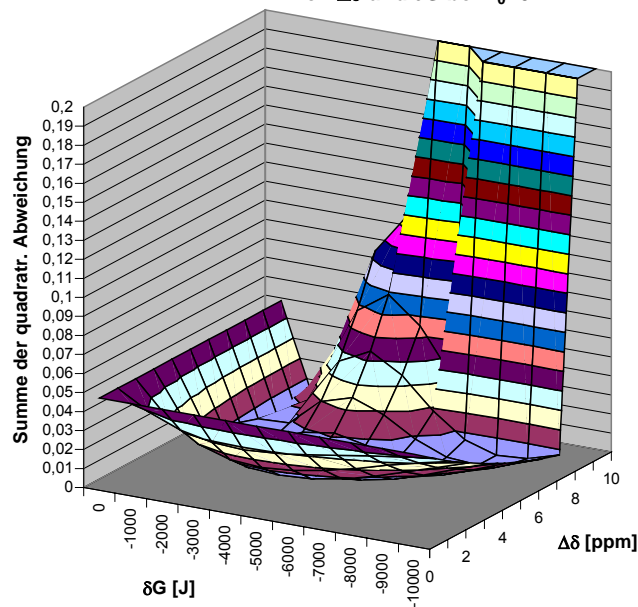
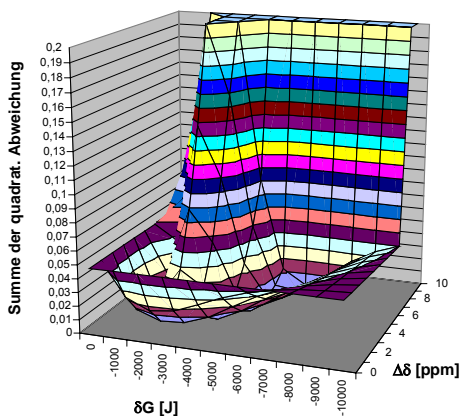


Abbildung 100: 3D- Plot über die Veränderung der Summe der quadratischen Abweichung bei unterschiedlichen Kombinationen aus  $\Delta G_0$  und  $\Delta\delta$  und  $T_0=0K$ . Die Farben illustrieren den Wert der Summe der quadratischen Abweichung. Man sieht einen relativ breiten Bereich mit gleich guten Abweichungen. Dieser Bereich wird bei Erhöhung von  $T_0$  sukzessive schmaler und enger.

Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\Delta\delta$  und  $\delta G$  bei  $T_0=135K$



Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\Delta\delta$  und  $\delta G$  bei  $T_0=200K$

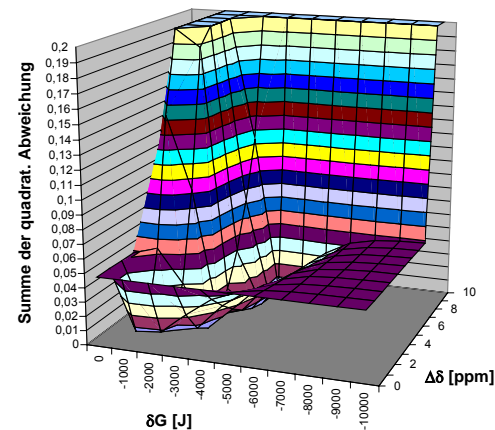
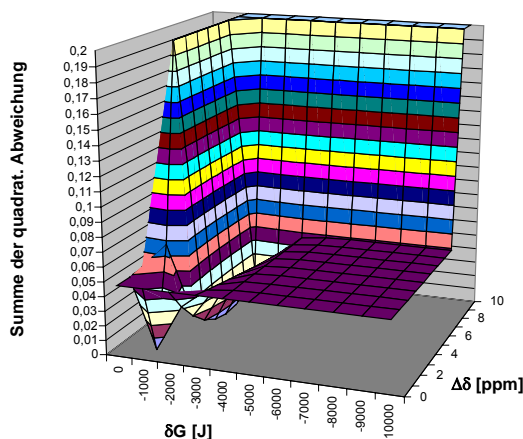


Abbildung 101: 3D- Plot über die Veränderung der Summe der quadratischen Abweichung bei unterschiedlichen Kombinationen aus  $\Delta G_0$  und  $\Delta\delta$  und  $T_0=135K$  (links), bzw.  $T_0=200K$  (rechts). Der anfangs große Bereich geringer Abweichung beginnt schmaler zu werden.

Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\Delta\delta$  und  $\delta G$  bei  $T_0=260K$



Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\Delta\delta$  und  $\delta G$  bei  $T_0=270K$

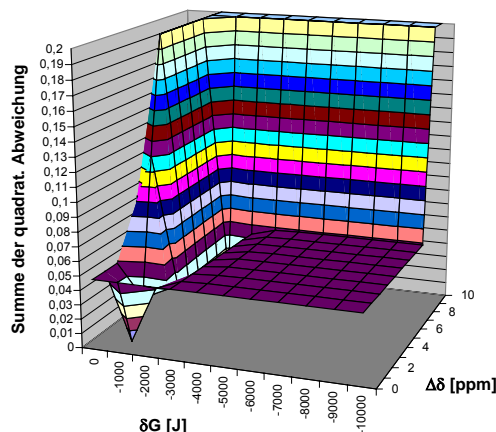


Abbildung 102: 3D- Plot über die Veränderung der Summe der quadratischen Abweichung bei unterschiedlichen Kombinationen aus  $\Delta G_0$  und  $\Delta\delta$  und  $T_0=260K$  (links), bzw.  $T_0=270K$  (rechts). Der Bereich wird zur Spalte und schließlich beschränkt sich der Bereich der geringsten Abweichung auf ein einziges definiertes Gebiet.

Die durchgeführten Untersuchungen zeigen, dass bei einem Wert von  $T_0$  im physiologischen Bereich (ca. 270- 280K) das Areal für die niedrigsten Werte sehr scharf definiert ist, während es bei einem  $T_0$  bei 0K noch einen großen Bereich gleich oder ähnlich guter Lösungen für die Fitfunktion gibt. Nun ist die Tatsache, dass in einem Bereich die Lösungen besser clustern als in einem anderen Bereich, zwar ein Indiz, aber noch kein Beweis für die Richtigkeit der Annahme. Deshalb haben wir in einem weiteren Test die Absolutwerte der Summe der quadratischen Abweichung bei Temperaturen von  $T_0=0K$  und  $T_0=273K$  untersucht. Leider sind die vorhandenen Daten der chemischen Verschiebungen zwar sehr gut, aber doch nicht völlig rauschfrei, weswegen, um bei dem Bild der eben dargestellten Grafiken zu bleiben, der Bereich der minimalen quadratischen Abweichungen immer auch Wellenberge und -täler aufweist. Die Folge davon ist, untersucht man die Absolutwerte der quadratischen Abweichungen, dass es bei einem Vergleich statistisch geringfügige Unterschiede geben muss, sodass zufällig einmal das absolute Minimum bei einem  $T_0$  von 0K und das andere Mal bei einem  $T_0$  von 273K gefunden wird. Dieses wiederum bedeutet, dass eine Aussage nur dann getroffen werden kann, wenn die Unterschiede genügend groß sind und signifikant auf ein bestimmtes  $T_0$  hindeuten.

#### Untersuchung der Summe der quadratischen Abweichung für $T_0=0K$ und $T_0=273K$

Unter diesen Rahmenbedingungen haben wir Tests durchgeführt, die für nahezu alle vorhandenen  $H^N$ - Atome Fits mit oben dargestellter Formel bei  $T_0$  von 0K und  $T_0$  von 273K generierten. Danach wurden die gefundenen minimalen Summen der quadratischen Abweichungen miteinander verglichen. Abbildung 103 zeigt die Differenz dieser Werte.

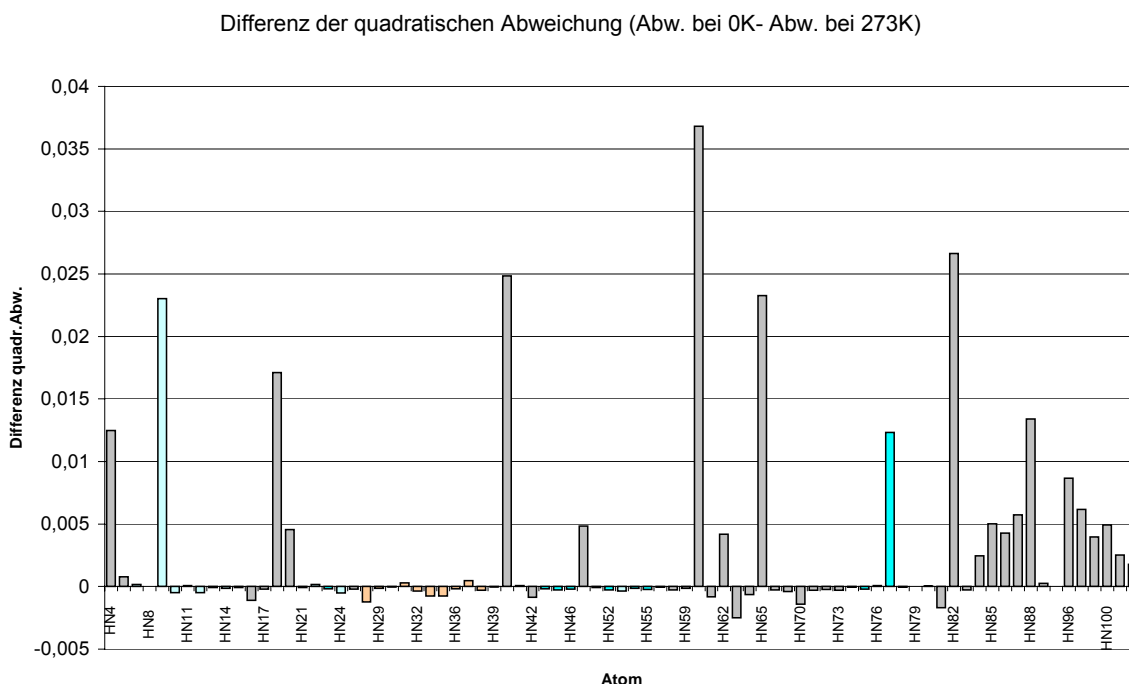


Abbildung 103: Differenz der Summe der quadratischen Abweichung bei einem Fit mit  $T_0=0K$  und  $T_0=273K$  ( $FQS(0K) - FQS(273K)$ ). Sekundärstrukturelemente sind farblich hervorgehoben. Man sieht bei deutlichen Unterschieden in den Differenzen, dass der Fit mit einem  $T_0$  bei 273K einen niedrigeren Wert erbringt

Man erkennt für alle Fälle, in denen es signifikante Unterschiede in der Summe der quadratischen Abweichungen gibt, dass der Fit mit einem  $T_0$  bei 273K einen besseren Wert erzielt. Für Fälle in denen der Unterschied in den quadratischen Abweichungen geringer ist als ca. 0,003 ist keine Aussage möglich, da dies dem lokalen Rauschen entspricht, das durch die experimentellen Daten und der Annäherung durch einen Fitalgorithmus entsteht.

Es sei jedoch hinzugefügt, dass die Atome mit signifikanten Unterschieden generell Atome darstellen, deren chemische Verschiebungen eine relativ starke Änderung im Laufe der Temperaturerhöhung aufweisen.

Ein weiteres Ergebnis dieser Untersuchungen und ein Indiz dafür, dass ein  $T_0$  bei 0K nicht geeignet erscheint, ist der Wert von  $(\delta_A - \delta_B)$ ,  $\Delta\delta$ . Dieser muss in der Gleichung für  $T_0=0K$  durchschnittlich Größenordnungen von über 7ppm annehmen (vgl. Abbildung 104), damit der Fit gelingt. Es erscheint aber relativ unwahrscheinlich, dass sich die chemischen Verschiebungen von  $H^N$ -Atomen durch das Öffnen, bzw. Schließen von Wasserstoffbrückenbindungen um einen derart großen Wert ändern, besonders wenn man in Betracht zieht, dass generell alle beobachteten  $H^N$ -Shifts im Bereich von ca. 8 bis 11ppm liegen.

Auch Williamson hat in ähnlichen Arbeiten [Wil03] ein  $\Delta\delta$  von 0,5 ppm angenommen, was sich gut mit den von uns gefundenen Werten bei einem  $T_0$  von 273K deckt.

Temperatur $T_0$ [K]	0	273
Mittlerer $\Delta\delta$ [ppm]	7,44	0,9
Mittleres $\Delta G_0$ [J]	-4650	-573
Mittlere Quadr. Abw.	0,0064195	0,00359127

Abbildung 104: Beobachtete Mittelwerte für  $\Delta\delta$ ,  $\Delta G_0$  und die Summe der quadratischen Abweichung. Letztere ist geringfügig höher bei einem  $T_0$  von 0K, allerdings ist hier der Unterschied nicht signifikant. Interessant ist der sehr hohe Wert für  $\Delta\delta$  bei  $T_0=0K$ . Insgesamt auffällig sind auch die relativ kleinen Werte für  $\Delta G_0$ , die z.B. bei  $T_0=273K$  nur rund 600J betragen.

Abbildung 105 zeigt beispielhaft das Verhalten der Summe der quadratischen Abweichung bei Anwenden der Fitfunktion auf ein Atom, das deutliche Unterschiede aufweist, je nachdem ob man ein  $T_0$  von 0K oder ein  $T_0$  von 273K in der Formel zugrunde legt.

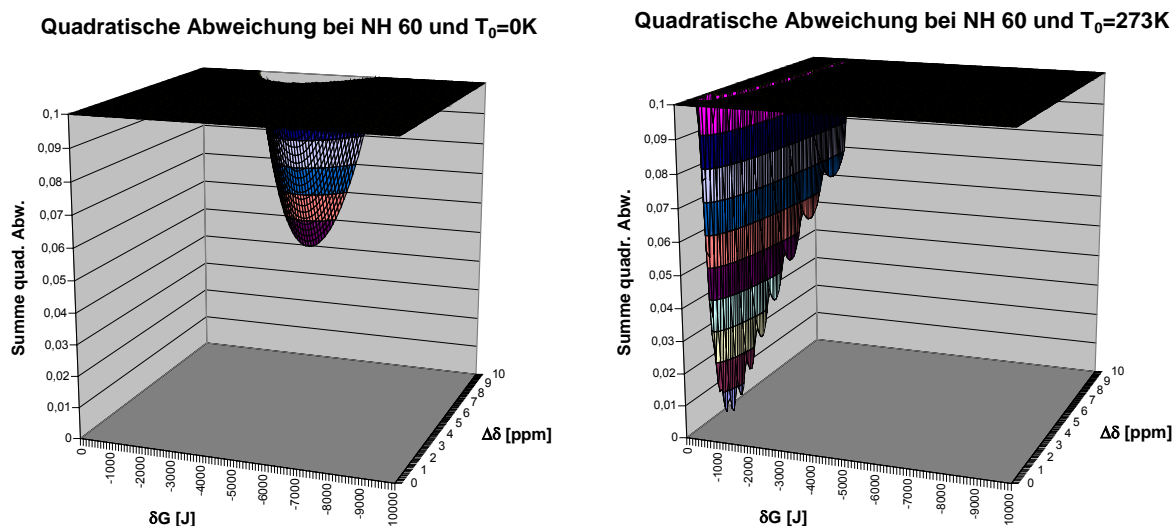
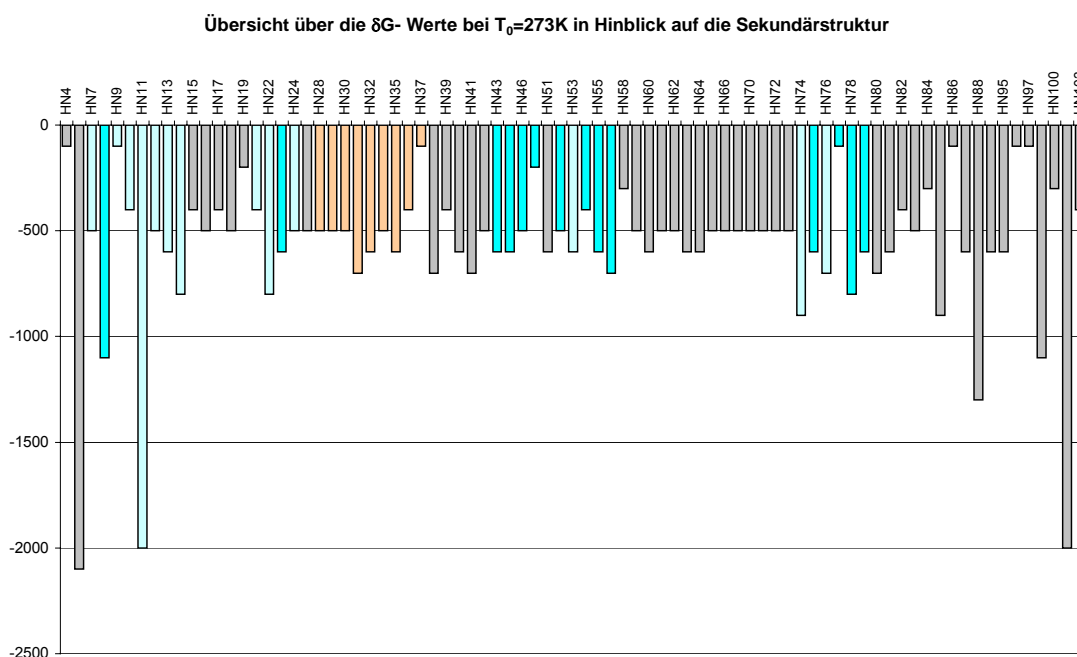


Abbildung 105: Darstellung der Summe der quadratischen Abweichungen bei  $H^N$  60. Links:  $T_0=0K$ , rechts:  $T_0=273K$ . Es ist klar zu sehen, dass die Werte bei  $T_0=273$  deutlich besser sind. Außerdem sind auch die unterschiedlichen  $\Delta G_0$  und  $\Delta\delta$ -Kombinationen zu erkennen, die sich für die beobachteten jeweiligen Minima ergeben.

Unterzieht man nun aber die Werte der freien Enthalpie einer näheren Untersuchung, so sind bei den Fits mit  $T_0=273K$  sehr kleine Werte zu beobachten. Allerdings steht eine genaue Erkenntnis darüber, was diese Beobachtungen bedeuten noch aus. Geht man davon aus, dass  $\Delta G_0$  die Energie repräsentieren sollte, die bei der Bildung einer Wasserstoffbrückenbindung frei wird, sind diese Werte zu gering ( $\Delta G_0$  bei H-O-Brückenbindung in Wasser ca.  $-6kJ$ ). Denkbar wäre es, dass das beobachtete  $\Delta G_0$  in den Untersuchungen über den Temperaturbereich nicht das vollständige Schließen, bzw. Lösen einer HO-Brückenbindung anzeigt, sondern lediglich kleinere Änderungen und



Umstrukturierungen. Dies stünde auch im Einklang damit, dass während der Messungen weder Kälte- noch Hitzedenaturation stattgefunden hat, sondern sich das Protein ständig im wohlgefalteten Bereich bewegte. Dabei kann man davon ausgehen, dass während des ganzen beobachteten Zeitraumes sich keine vorhandene Bindung komplett gelöst hat. Auch sind keine Unterschiede in den Energiewerten zu erkennen, die im Zusammenhang mit Sekundärstrukturelementen stehen könnten (vgl. Abbildung 106).



*Abbildung 106: Auftragung der  $\Delta G_0$ -Werte bei einem fit der Daten mit  $T_0=273K$ . Sekundärstrukturelemente sind farblich hervorgehoben, Faltblatt blau, Helix rot. Experimentell bestimmte Wasserstoffbrücken sind durch dunklere Farbgebung gekennzeichnet. Die Betrachtung der Energien weist keinen Zusammenhang mit Sekundärstrukturelementen oder experimentell bestimmten HO-Brücken auf.*

Betrachtet man jedoch die  $\Delta G_0$ -Werte bei den Fits mit  $T_0=0K$ , so sind diese deutlich höher und teilweise durchaus in der Größenordnung der Bindungsenergie einer Wasserstoffbrückenbindung (Abbildung 107). Bei näherer Prüfung könnte man auch leicht erhöhte Energiewerte in den wohlgefalteten Bereichen der Sekundärstrukturelemente postulieren, was darauf schließen ließe, dass diese, wie allgemein bekannt, durch Wasserstoffbrücken stabilisiert und gebunden sind. Geht man nun davon aus, dass man mittels des Verwendens von  $T_0=0K$  einen viel größeren Bereich „untersucht“, bzw. dafür eine mathematische Vorhersage trifft, so schließt das naturgemäß auch den Punkt mit ein, an dem alle potentiellen HO-Brücken fest geschlossen und gebunden sind, und den Punkt, an dem alle geöffnet vorliegen. Dann wiederum könnte das berechnete  $\Delta G_0$  durchaus die Energie repräsentieren, die bei dem Schließen einer HO-Brücke frei wird.



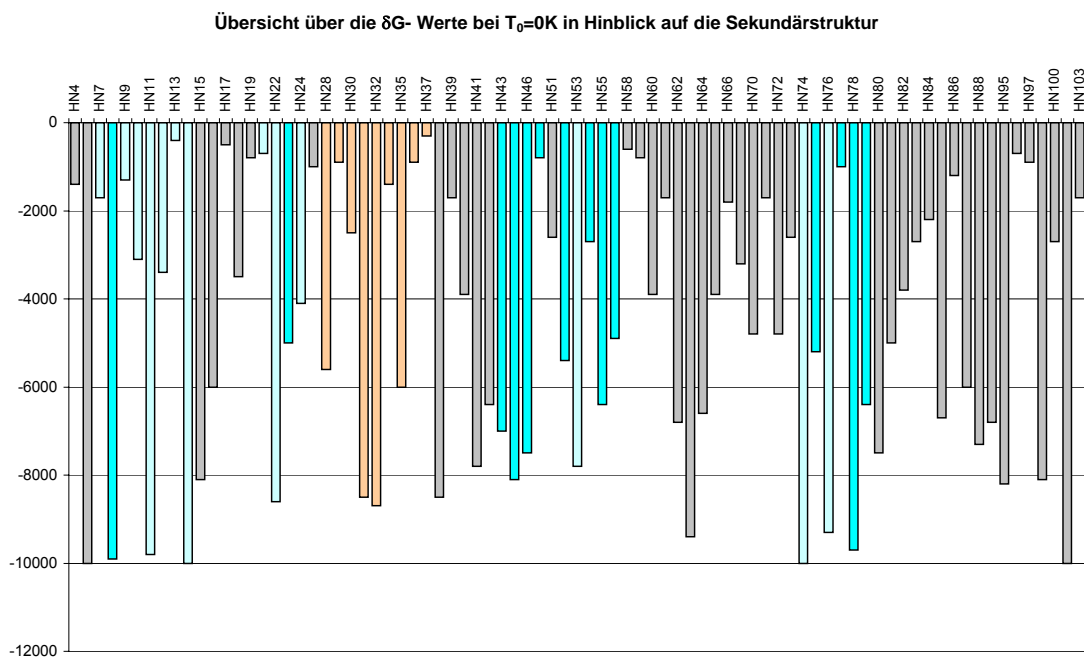


Abbildung 107: Auftragung der  $\Delta G_0$ -Werte bei einem fit der Daten mit  $T_0=0K$ . Sekundärstrukturelemente sind farblich analog Abbildung 106 hervorgehoben. Die Betrachtung der Energien in diesem Diagramm weist eine deutlich ungleichmäßigere Verteilung der Werte zwischen ca. 500J und 10000J auf. Jedoch sind auch hier eindeutige Zusammenhänge von  $\Delta G_0$ -Werten und Sekundärstruktur kaum zu erkennen.

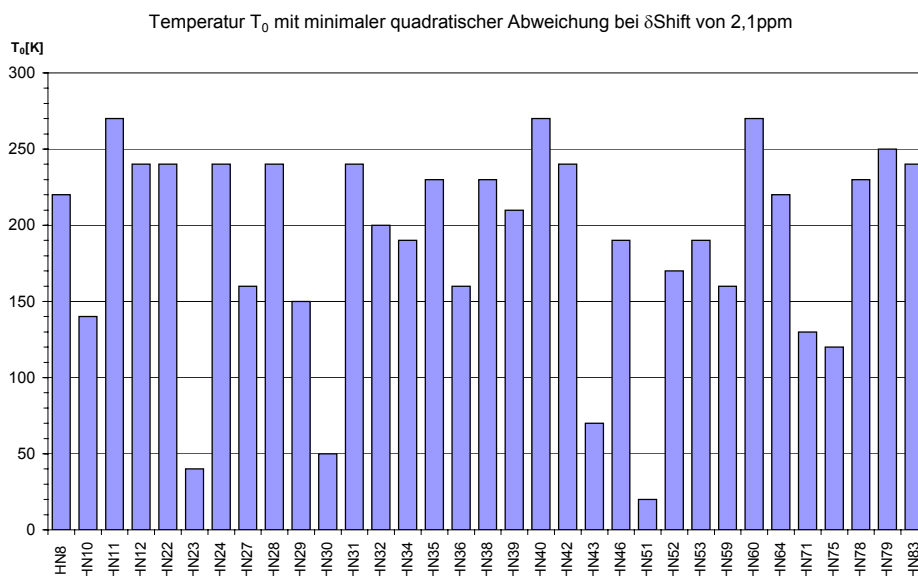
Nun sind diese Ausführungen allerdings rein mathematisch und theoretisch, da vor dem Erreichen der absoluten Bewegungsunfähigkeit bei Proteinen die Kältedenaturation einsetzt, und spätestens dann undefinierbare und große Änderungen in den chemischen Verschiebungen der Atome zu erwarten sind, genau wie am anderen Ende der Skala, bei der Hitzedenaturation. Allein diese Tatsache macht eine Vorhersage über diesen Bereich hinaus unsicher und unmöglich.

Vor diesem Hintergrund betrachtet erscheint es sinnvoll, in Betracht zu ziehen, dass  $T_0$  den Punkt der Kältedenaturation bezeichnen könnte, an dem in vivo die Bewegungsfähigkeit ihr Minimum erreicht, bevor die räumliche Struktur verloren geht. In diesem Zusammenhang wären weitere Untersuchungen anzustreben, die Messdaten bis hin zu Kälte- und Hitzedenaturation liefern.

### Ausblick, weitere Arbeiten

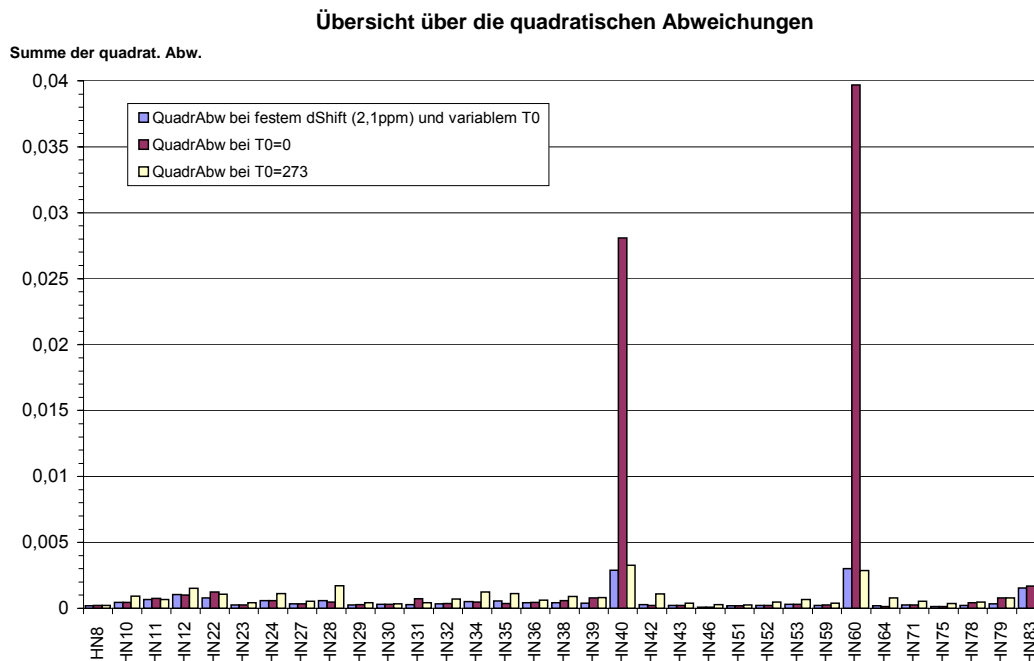
In ersten Untersuchungen dazu haben wir in Anlehnung an die Untersuchungen von Wagner et al. [Wag83], die für den Unterschied der chemischen Verschiebungen für eine völlig geöffnete und eine geschlossene Wasserstoffbrücke den Wert von 2,1ppm bestimmt haben, ein Fit der Daten mit freiem  $T_0$  und dem Wert von 2,1ppm für  $\Delta\delta$  versucht.

Erste Ergebnisse mit Fits an einigen zufällig ausgewählten Atomen zeigen die besten Werte der Summe der quadratischen Abweichung bei  $T_0$ -Werten weit über 0K, aber deutlich unter 273K, bei durchschnittlich ca. 200K. Natürlich ist auch hier ein gewisses Rauschlevel vorherzusehen, das für die Bestimmung von  $T_0$  nicht für alle Atome gleiche Werte erwarten lässt.



*Abbildung 108: Fit der Daten der chemischen Verschiebungen bei freiem  $T_0$  und festgelegtem  $\Delta\delta$ . Dargestellt ist das jeweils für den besten Fall berechnete  $T_0$ . Man beobachtet in Extremfällen starke Schwankungen von ca. 50K bis 270K. Die meisten berechneten  $T_0$  Temperaturen bewegen sich aber im Bereich von 150 bis 200K.*

Betrachtet man auch die Absolutwerte der Summe der quadratischen Abweichungen, so sind die Werte in allen Fällen bei dem Fit mit freiem  $T_0$  und einem  $\Delta\delta$  - Wert von 2,1ppm am niedrigsten. Allerdings sind die Unterschiede so gering, dass sie auch, wie oben schon, auf das Rauschen der Daten zurückzuführen sein könnten.



*Abbildung 109: Übersicht über die Absolutwerte der Summe der quadratischen Abweichungen. In nahezu allen Fällen ist dieser Wert bei einem Fit der Daten mit freiem  $T_0$  und  $\Delta\delta$  von 2,1ppm am kleinsten. Allerdings sind die beobachtbaren Unterschiede sehr gering.*

Zusammenfassend ist aufgrund aller gewonnenen Erkenntnisse zu beachten, dass der Effekt, den es hier zu beobachten gilt, sehr klein gegen das vorhandene Rauschen der Daten ist. Dieses wiederum erschwert letztlich die exakte Quantifizierung und Beobachtung des gewünschten Effektes. Daraus resultieren auch die Schwankungen in den Absolutwerten der Summe der quadratischen Abweichung (vgl. Anmerkungen zu Abbildung 103), was bedeutet, dass eine Aussage nur bei deutlichen Unterschieden getroffen werden kann, was wiederum nur für einige wenige Atome zutrifft.

Anzumerken ist weiterhin, dass die Shiftänderungen nicht allein von fehlenden oder gebildeten Wasserstoffbrückenbindungen hervorgerufen werden, sondern auch kleine lokale Veränderungen in der Struktur den Einfluss von OH- Bindungen überlagern.

Nicht zu verachten ist auch das Problem des Fits der Daten mit vorstehender Formel, die für ein eindeutiges Ergebnis zu viele Freiheitsgrade aufweist, sodass Parameter vorgegeben werden müssen (z.B.  $T_0$  oder  $\Delta\delta$ ). Auch dann noch sind diese Berechnungen sehr rechenintensiv und zeitaufwendig.

Für detailliertere Aussagen in diesem Bereich müssten entweder einige Parameter experimentell bestimmt (z.B. experimentelle Bestimmung des Kältedenaturationspunktes) und/oder eine Erweiterung des Daten- und Messbereiches angestrebt werden. Messungen bis hin zu Kälte- und Hitzedenaturation könnten Aufschluss darüber geben, wie sich die chemischen Verschiebungen an den Randbereichen verhalten, und welches mathematische Modell, bzw. welches  $T_0$  verwendet werden soll.

Diese Messungen sind allerdings momentan wegen technischer Gegebenheiten nicht möglich, da mit dem vorhandenen Kryo- Probenkopf zwar sehr rauscharme und signalstarke Spektren aufgenommen werden können, allerdings wegen des herrschenden Temperaturunterschiedes zwischen Probe und Messkopf keine Temperaturen über 60 °C und unter 5°C vermessen werden können.

Nichtsdestoweniger haben die gezeigten Untersuchungen die Hoffnungen darin bestärkt, dass das Temperaturverhalten der chemischen Verschiebungen mathematisch vorhergesagt werden kann. Außerdem zeigten sich deutliche Abweichungen des Shiftverhaltens von der Linearität, was zwar seit längerem postuliert, aber hier erstmals an einer umfangreichen Messserie experimentell bestätigt wurde.

## 5.5. Ausblick

Weitere Untersuchungen in diesem Gebiet, die darauf abzielen Strukturänderungen an den Punkten der Kälte- und Hitzedenaturation zu sehen, müssten mit dem Versuch einhergehen, Messreihen über 50°C und unter 10°C durchzuführen. Da das Protein Saratin bis etwa 75°C stabil und wohlgefaltet ist, müsste angestrebt werden diesen Punkt bei NMR- Messungen zu überschreiten. Wie bereits ausgeführt, sind die technischen Möglichkeiten dazu momentan beschränkt.

Im Bereich der mathematischen Beschreibung der Shiftänderungen könnte mit experimentellen Methoden der Kältedenaturationspunkt bestimmt, und dieser als  $T_0$  in die besprochene Gleichung eingesetzt werden. Es wird sich zeigen, ob der ermittelte Kältedenaturationspunkt mit dem berechneten Punkt des  $T_0$  bei einer maximalen Änderung der chemischen Verschiebung von 2,1ppm in Einklang zu bringen ist. Untersuchungen und Tests dazu sind bereits in Vorbereitung.

Des weiteren ist geplant, das Temperaturverhalten aller vorhandenen NOEs in dem gemessenen Bereich von 10°- 50°C zu analysieren. Die Daten dazu sind dank der Funktionsweise des zur Zuordnung verwendeten Programms AUNOAS bereits vorhanden bedürfen aber noch der Auswertung. Die Untersuchungen speziell an z.B. den  $H_\alpha$ - Shifts könnten Aufschluss darüber geben, ob eventuell vorhandene strukturelle Änderungen am Verhalten der  $H^N$ - Shifts beteiligt sind. Dadurch könnte eine Entkopplung von Shiftänderungen durch Öffnen von Wasserstoffbrückenbindungen und Strukturänderung erreicht werden.

Das Fernziel aller dieser Bemühungen ist es natürlich wie in allen Fällen, ein tieferes Verständnis der Strukturdynamik zu erreichen. Dazu wäre auch ein Vergleich des Temperaturverhaltens mit dem Verhalten des Proteins bei erhöhtem Druck sicherlich nützlich und denkbar. Dieser Intention kommt zugute, dass an unserem Institut ausgereifte Technologien zur Messung unter Hochdruckbedingungen verfügbar sind. Auch hierzu sind schon Gespräche über durchzuführende Tests erfolgt, und erste Versuche sollen noch in diesem Jahr begonnen werden.

## 6. Diskussion und Ausblick

Für die Bestimmung von Proteinstrukturen ist es notwendig 2D oder 3D NOESY Spektren auszuwerten und die dort auftretenden Peaks (NOEs) so vollständig, und vor allem so richtig wie möglich zu benennen. Dieser Prozess allerdings ist sowohl sehr zeitaufwendig, als auch kompliziert und bislang nur in Handarbeit möglich, da viele unterschiedliche Randbedingungen bei der Auswertung von NMR- Spektren miteinbezogen werden müssen. Deshalb werden und wurden weithin Anstrengungen unternommen, die Zuordnung der Resonanzen zu den einzelnen Spinsystemen mittels geeigneter Algorithmen computergestützt ablaufen zu lassen, um somit den Zeitrahmen für die Erzeugung einer NMR- Struktur von momentan etwa 1 bis 2 Jahren zu verkürzen, und damit die Konkurrenzfähigkeit zu der zweithäufigsten Methode zur Aufklärung von Proteinstrukturen, der Röntgenkristallographie, zu erhöhen.

Ein besonderes Problem bei der Zuordnung stellen die Signale von diastereotopen Protonen der Seitenketten der einzelnen Aminosäurereste dar. Bei diesen Paaren von Protonen gestaltet sich die Zuordnung der Resonanzen zu ihren entsprechenden Signalen (im Folgenden als stereospezifische Zuordnung bezeichnet) in der Praxis oftmals recht zeitaufwendig und schwierig, da nicht mit einfachen Mitteln geklärt werden kann, welche Resonanz zu welchem Proton gehört. Dazu müssen die J- Kopplungen, die sich auf die dihedralen Winkel der Seitenkette beziehen, bestimmt, und mit der Karplus- Gleichung die jeweiligen Winkel berechnet werden. Zusätzliche Schwierigkeiten treten durch die häufigen Überlappungen der Resonanzsignale bei größeren Biomolekülen auf. Da jedoch Strukturrechnungen auf der Basis von Abständen, besonders weitreichenden Abständen, basieren, ist es wichtig, diese Zuordnungen richtig zu treffen.

Um diese Aufgabe zu erleichtern haben wir das im ersten Teil der Arbeit vorgestellte Programm, STEREO entwickelt. Es ist ein einfach zu bedienendes Werkzeug, das auf dem Vergleich von experimentellen und simulierten  $^1\text{H}$  NOESY NMR Spektren basiert. Bei diesem Vergleich werden zusammengehörende Methylenprotonen oder Atomgruppen von Seitenketten der gleichen Aminosäure, wenn sie nicht unterscheidbar sind, identifiziert und ihre korrekte Benennung mittels Berechnung eines NMR R- Faktors überprüft. Um die Richtigkeit solcher Zuordnungen zu überprüfen, wurde, wie bereits an anderer Stelle [Gro00, Kir01] beschrieben, ein Computerprogramm (RFAC) entwickelt, welches die automatische Beurteilung von dreidimensionalen NMR- Proteinstrukturen erlaubt, und durch die Berechnung eines R- Faktors ein zuverlässiges Kriterium für die Qualität dieser Strukturen bietet.

STEREO wurde erfolgreich an drei mittelgroßen Proteinen getestet: dem Kälteschockprotein (*TmCSP*) aus *Thermotoga maritima*, dem Histidine containing Phosphorcarrier Protein (HPr) aus *Staphylococcus carnosus*, und der RAS- Bindenden Domäne des RAL Guanosin Nucleotid dissociation stimulators (RalGDS) aus *Homo sapiens*. Die Ergebnisse zeigen, dass eine korrekte stereospezifische Zuordnung den R- Wert verringert. Auch ist es ziemlich aufschlussreich, ein Bündel von NMR- Strukturen zu untersuchen. Dadurch können

statistisch gesicherte Aussagen darüber getroffen werden, ob man in der Zuordnungsliste die Namen eines Paares stereospezifischer Atome tauschen soll, oder nicht. Abhängig von der Anzahl der Strukturen in welchen diese Atome als falsch zugeordnet klassifiziert wurde, ist es möglich, darüber zu entscheiden.

Bei den diversen Tests an unterschiedlichen NMR- Spektren haben wir einen gewissen Zusammenhang zwischen der Benennung eines stereospezifischen Signals und seiner chemischen Verschiebung (in ppm) beobachtet. Es stellte sich heraus, dass es eine Korrelation zwischen einer kleineren chemischen Verschiebung und der Bezeichnung mit der kleineren Nummer, z.B.  $H^{\beta 3}$ , gibt. Um diese beobachtete Beziehung noch näher zu untersuchen, haben wir Datenbankeinträge statistisch ausgewertet [Bmr01]. Nach einem Fit der Daten mit einer zusammengesetzten Gauß- Funktion zeigte sich, dass in den meisten der beobachteten Fälle, tatsächlich der kleinere Shift zu der Benennung als z.B.  $H^{\beta 3}$  gehört.

Zur weiteren Verifizierung der bereits gewonnenen Erkenntnisse ist geplant, zusätzliche Daten auszuwerten, bei denen die stereospezifische Zuordnung mit herkömmlichen Methoden durchgeführt wurden. Des weiteren sollen in Zukunft, zusätzlich zu der Berechnung von R- Werten, Bayes- Wahrscheinlichkeitskriterien eingeführt werden, die sich aus der oben erwähnten Auswertung von 1668 in der BMRB- Datenbank vorhandenen  $^1H$  chemischen Shift- Einträge und ihrer Überprüfung auf die unterschiedlichen stereospezifisch zugeordneten Signale ergeben. So lässt sich beispielsweise bei genügender Datenmenge eine statistische Aussage darüber treffen, welches Signal in welcher Aminosäure welche chemische Verschiebung einnimmt, genauer gesagt, ob in einer gegebenen Aminosäure z.B. das  $H^{\beta 2}$  den niedrigeren oder höheren Wert der chemischen Verschiebung aufweist, als das  $H^{\beta 3}$ . Dazu wurden bereits Vorarbeiten dahingehend geleistet, dass die Daten erfasst und ausgewertet wurden. Es wurde für jede Aminosäure eine Wahrscheinlichkeitsverteilung ihrer stereospezifischen Gruppen erstellt und in eine mathematische Formel gefasst.

Zusammenfassend lässt sich sagen, dass STERO ein nützliches Hilfsmittel darstellt, die stereospezifische Zuordnung bei einer Proteinstruktur schnell und einfach zu überprüfen und gegebenenfalls zu korrigieren. Dies stellt einen weiteren kleinen Schritt dar, die Qualitätsansprüche, die heutzutage an eine Proteinstruktur gestellt werden, zu erfüllen, die Strukturen zu verbessern und gleichzeitig den Arbeitsaufwand zu verringern.

Indes ist nach wie vor unbestritten, dass der Arbeitsaufwand, der zu leisten ist, um die vielen tausend Peaks in der Zuordnung zu suchen, zu identifizieren und schließlich und endlich für die Strukturrechnung verfügbar zu machen, sehr groß ist, und den Hauptteil der Bemühungen einer Person ausmacht, die eine räumliche NMR- Proteinstruktur erstellen will. Gleichzeitig ist dies ein Prozess, der bisher immer noch von Hand durchgeführt werden musste und daher den limitierenden Schritt bei der Erstellung einer Struktur darstellt.

Im Bereich der NMR werden Atome einer Proteinsequenz anhand ihrer chemischen Verschiebung innerhalb der Spektren identifiziert. Am Anfang der Strukturaufklärung werden die zu den einzelnen Atomen gehörenden Shifts ermittelt und in der sogenannten sequenziellen Zuordnung zusammengefasst. Das Problem, das sich dabei u. a. ergibt ist, dass die Shifts der verschiedenen Atome aus der sequentiellen Zuordnung nicht exakt zu den in weiteren NOESY Spektren desselben Proteins gefundenen chemischen Verschiebungen

passen, da die chemische Verschiebung eines Atoms auch von Parametern wie Temperatur, Druck, Salzgehalt der Probe u.a. abhängt. Deshalb können Zuordnungen nur bedingt auf Spektren übertragen werden, die unter leicht abweichenden Bedingungen aufgenommen wurden. Dies führt vor allem in Bereichen, in denen die Signale sehr dicht liegen, zu mehreren Möglichkeiten der Zuordnung. Die damit verbundenen Unsicherheiten pflanzen sich in die Strukturbestimmung fort, und führen zu schlecht definierten Strukturen. Weiter ist es aus den eben genannten Gründen momentan noch nicht möglich zu verfolgen, wie sich z.B. das strukturelle Verhalten eines Proteins ändert, wenn äußere Parameter variiert werden.

Das von uns entwickelte Programm AUNOAS (AUTomated NOe ASSignment) stellt eine Lösung für diese Probleme bereit, indem es mittels eines iterativen, auf sich selbst rückkoppelnden Algorithmus sowohl das betrachtete NMR- Spektrum zuordnet, als auch die vorhandene sequenzielle Zuordnung optimal an das eben verwendete experimentelle Spektrum anpasst.

Der Algorithmus, der dem Programm AUNOAS zugrunde liegt, stellt im Wesentlichen die „menschliche Vorgehensweise“ nach und versucht, von sicheren Erkenntnissen auf noch „unsichere“ Erkenntnisse zu schließen. Dieser Prozess wird iterativ wieder auf sich selbst rückgekoppelt und optimiert so ständig seine Eingabeparameter. Wertet man von Hand ein NOESY Spektrum aus, so beginnt man in der Regel mit isoliert liegenden NOEs und versucht diese eindeutig zu benennen. Im nächsten Schritt untersucht man Peaks die noch halbwegs gut von ihren Nachbarn getrennt sind. Wenn es für diese Signale eindeutige Zuordnungsmöglichkeiten gibt, so notiert man die daraus gewonnenen Erkenntnisse, und versucht damit im weiteren Verlauf überlappende Peaks zu benennen und eventuell noch Strukturinformation aus stark überlappten Bereichen zu ziehen.

Dieser iterative, rückgekoppelte Optimierungsprozess sorgt dafür, dass das Programm gut mit Daten arbeiten kann die entweder verrauscht sind, oder von den Erwartungswerten der anfänglichen Zuordnung abweichen. Bei der Programmarchitektur wurde großer Wert auf Flexibilität in der Anwendung gelegt. D.h. es ist durch geeignete Wahl der Funktionsparameter möglich, sowohl Spektren auszuwerten für die noch keine, oder nur eine unsichere Strukturvorstellung existiert, als auch das Verhalten von Proteinen, deren Struktur bekannt ist, unter z.B. Veränderungen der äußeren Bedingungen zu analysieren.

Zur Überprüfung der richtigen Funktionsweise haben wir das Programm zunächst mit synthetischen, überprüfbaren Daten getestet.

Ein Spektrum des Kälteschockproteins *TmCSP* wurde mittels pdb- und meta- File rückgerechnet. Das ergab eine Masterliste mit ca. 6000 benannten Peaks. Bei einem Spektrum mit 6000 Signalen liegen diese sehr dicht, außerdem ist eine große Zahl an überlappenden Peaks zu erwarten. Diese Masterliste wurde AUNOAS als experimentelle, unbenannte Masterliste übergeben und zugeordnet.

Nach dem Programmlauf wurde überprüft, wie viele Signale richtig zugeordnet worden waren, abhängig von für den Programmlauf gewählten Fangradien. Es zeigte sich, dass auch für extrem große Radien von 0,2 ppm die Fehlerquote sehr gering ausfiel (ca. 1%).



Daraufhin wurde das Spektrum sowohl in der w1 als auch in der w2 Koordinate mittels einer Gauß- Funktion unterschiedlicher Halbwertsbreiten verrauscht. Dies simulierte ein experimentelles Spektrum, bei welchen die Signale eines Spinsystems nicht exakt auf einer Koordinate zu finden sind. Auch hier arbeitete das Programm zuverlässig und in vernünftigen Grenzen fehlerfrei. Um auch andere, in experimentellen Spektren auftretende Störungen und ihre Auswirkungen auf das Programm zu untersuchen, wurden anschließend von uns unterschiedliche Störungen simuliert.

Zunächst der „Störungstyp“ der Verunreinigung, was bedeutet, dass neben dem zu messenden Proteinsignalen äquivalente Signale existieren, die z.B. von teilweise denaturiertem Protein oder von vorhandenem Fremdproteinen oder anderen Zusätzen stammen können. Um dieses zu untersuchen, wurden Signale hinzugefügt, die gleich hohe Volumen wie die eigentlichen Signale aufwiesen, und auch einen guten Qualitätswert von AUREMOL zugewiesen bekamen, da diese Signale ja auch wirkliche Signale sind.

Als zweites wurden Störungen hinzugefügt, die durch Artefakte zustande kommen. Artefakte entstehen durch die Aufnahme und Prozessierung von Spektren und weisen gleiche Volumen wie die Signale auf, werden aber aufgrund ihrer eigenen anderen Peakform und Symmetrie von AUREMOL mit niedrigen Qualitätswerten bedacht.

Zuletzt sollte noch Rauschen simuliert werden, das in experimentellen Spektren allgegenwärtig ist, und den manuellen Zuordnungsprozess sehr erschwert. Rauschpeaks zeichnen sich durch ein niedrigeres Volumen und damit verbunden auch durch einen niedrigeren Qualitätswert aus. Allerdings können auch wirkliche Signale stark verringerte Volumen aufweisen, wenn diese Signale von sehr langreichweitigen Kontakten stammen. Dann wiederum sind diese Peaks von großer Bedeutung für die spätere Strukturbestimmung. Deshalb ist es sehr wichtig, dass ein Programm zur automatischen Zuordnung von Spektren auch möglichst genau zwischen Rauschen und Signal unterscheidet.

Wie sich zeigte, beeinflusst keiner der angebotenen Störungstypen den Programmablauf.

Egal, ob es sich um Verunreinigungen, Artefakte oder Rauschen handelt, und auch weitgehend unabhängig von der absoluten Anzahl der vorhandenen Störungen. In allen Fällen wurden die Signale des verwendeten Proteins richtig und vollständig zugeordnet. Mit zunehmender Anzahl der Störungen wurden bis maximal 18 Signale, die Störungen darstellen, benannt. Eine verschwindend geringe Zahl bei einem Signalaufkommen von dabei mehr als 10000 Peaks.

Nach diesen Tests haben wir ausgehend von den vorher beschriebenen Möglichkeiten zur automatischen Zuordnung von 2D NOESY- Spektren in einem umfangreichen Praxistest eine Temperaturreihe eines Proteins in Angriff genommen. Dazu wurde Saratin, ein Protein zur Inhibition der Blutgerinnung aus dem Blutegel verwendet. Dieses Protein ist über einen weiten Temperaturbereich (Hitzedenaturation bei ca. 75°C) stabil, liefert gute Signale, und war in ausreichender Menge am Lehrstuhl vorhanden. Außerdem existiert für Saratin eine nahezu vollständige sequenzielle Zuordnung und eine gute Vorstellung über die räumliche Struktur bei 298K. [Dom04] Mit diesem Protein haben wir in einer zweieinhalb Wochen



dauernden, unterbrechungsfreien, Messreihe 41 2D- NOESY- Spektren in einem Temperaturbereich von 10°C bis 50°C in 1° Schritten aufgenommen.

Bei der Berechnung der Strukturen kamen nur automatisch zugeordnete NOEs zur Verwendung, keine zusätzlichen experimentellen Daten, wie z.B. H- Brücken oder Winkelbeschränkungen. Da die Anzahl der von AUNOAS benannten Signale pro Spektrum zwischen ca. 2000 und 3000 schwankte, waren genügend Informationen für den Strukturrechenprozess vorhanden. (Zum Vergleich: manuell benannte 2D NOESYs bestehen aus ca. 1000 benannten Signalen). Nach der automatischen Zuordnung mit AUNOAS waren benannte Masterlisten (2D- Spektren) für alle untersuchten Temperaturen von 10°C bis 50°C verfügbar. Damit konnte für jede Temperatur eine räumliche Struktur von Saratin berechnet werden. Außerdem entstand für jede Temperatur eine an das jeweilige experimentelle Spektrum optimal angepasste sequenzielle Zuordnung. Diese Zuordnungen machten es zusätzlich möglich, das Verhalten der einzelnen chemischen Verschiebungen aller Atome abhängig von der Temperatur zu verfolgen.

Der Prozess von dem Beginn der Aufnahme der Daten, über die Prozessierung und Zuordnung, bis hin zur Berechnung der 41 Strukturen dauerte etwa 3 Monate, wovon auf das Aufnehmen der Spektren ca. 2 Wochen, auf das Prozessieren und Peakpicken 3 Wochen und das Berechnen der Strukturen ca. 4 Wochen entfielen. Die eigentlich zeitraubende und langwierige Arbeit der Zuordnung der Spektren - inklusive der Vorbereitung der Rechenfiles für das Programm- belief sich auf durchschnittlich 1 Stunde pro Spektrum. Dies bedeutet eine große Erleichterung und Zeitersparnis und ermöglicht auch neue, interessante, bisher wegen der langen Auswertdauer jedoch nicht praktikable Anwendungen, wie z.B. Temperatur- Druck oder pH- Reihen.

Die automatisch erzeugten Strukturen zeichneten sich allesamt durch gute RMSD- Werte und Gesamtenergien aus, was den Schluss zulässt, dass einerseits bei der automatischen Zuordnung keine Fehler entstanden sind, und andererseits das Aussehen und die Lagebeziehungen der Strukturelemente der Realität entspricht. Die Betrachtung der Strukturen zeigte, dass Saratin über den gesamten Temperaturbereich stabil ist (Hitzenaturierung erst ab mehrstündigem Erwärmen auf 75°C, Kältenaturierung nicht durchgeführt) und eine wohlgefaltete räumliche Struktur aufweist, die aus 5 antiparallelen  $\beta$ - Faltblättern und einer  $\alpha$ - Helix besteht. Es zeigte sich weiterhin, dass wie erwartet, die Loop- Bereiche beweglicher als die Sekundärstrukturelemente sind. Was nicht beobachtet werden konnte, war eine definierte Bewegungsrichtung einiger oder aller Teile der Struktur, stattdessen wurde eine gewisse Oszillation im Molekül sichtbar, die sich bei filmischer Betrachtung in einem „Pulsieren“ der 3D- Struktur mit der Temperatur äußert. Diese Tatsache gab Anlass zu näheren Untersuchungen. Für einen ersten Test haben wir die strukturellen Abstände von benachbarten CA- Atomen graphisch untersucht.

Wie man sehen konnte, weisen die Änderungen in den Abständen eine gewisse Regelmäßigkeit auf, und scheinen sich nicht statistisch zu verhalten, oder mit jeder Temperaturänderung willkürlich zu verändern. Es ist alle 3- 5 Temperaturschritte eine Erhöhung, bzw. Verringerung der betrachteten Abstände sichtbar, was auch auf eine definierte Periodizität schließen lässt. Die Vorstellung dazu, ist, dass auch bei wohlgefalteten

Molekülen unterschiedliche Energieniveaus (Feinstruktur) existieren, in die die Moleküle bei Änderung der Temperatur teilweise bis vollständig überführt werden.

Natürlich sind die untersuchten Daten nicht frei von Rauschen, was bedeutet, dass die Beobachtungen nicht beliebig genau durchgeführt werden können, genauso wenig, wie eine Aussage darüber möglich ist, wie die „genauen“ räumlichen Strukturen bei einer bestimmten Temperatur aussehen, da man in der NMR immer Mittelwerte beobachtet. Da aber die Oszillation deutlich über 1° Schritte hinausgeht, kann man in diesem Fall davon ausgehen, dass ein anderer Effekt daran beteiligt ist.

Um die dargestellten Vermutungen weiter zu sichern und zu untersuchen, haben wir es in Angriff genommen, auch CA- Atome weiter voneinander entfernter Aminosäuren zu betrachten. Bei diesen dürfte die beobachtete Oszillation noch stärker zutage treten, als bei benachbarten Atomen. Erste Versuche zeigten, dass diese Annahme korrekt ist, und gerade diese Atome eine relativ deutliche Periodizität aufweisen.

In einem weiteren Schritt sollen in Zukunft einerseits diese Untersuchungen auf alle Atome ausgedehnt werden, wobei unterschiedliche sequenzielle Abstände der Atome zu wählen sind, um die Oszillationen zu untersuchen. Andererseits sollte darüber hinaus im Anschluss eine mathematische Beschreibung der beobachteten Periodizität gefunden werden.

Dies dürfte zu einer Sicherung und einem besseren Verständnis der Strukturschwingungen führen.

Aus der Arbeitsweise von AUNOAS und den damit durchgeführten Zuordnungen ergaben sich 41 Listen von sequenziellen Zuordnungen für die untersuchten Temperaturen. Dadurch wurde es uns möglich, eine Betrachtung aller  $H^N$ - Shifts über den Temperaturbereich von 283K bis 323K in 1K- Schritten durchzuführen, und die Änderungen der chemischen Verschiebungen bei Änderung der Temperatur zu verfolgen.

Die Hoffnung bei einer genügend großen Anzahl von ausreichend guten Datenpunkten war es, eine Vorhersage mittels eines mathematischen Modells über das weitere Verhalten der chemischen Verschiebungen bei anderen Temperaturen treffen zu können. Dieses könnte für zukünftige Strukturbestimmungsbemühungen außerordentlich hilfreich sein.

Viele Kurven zeigten einen nicht linearen Verlauf, worauf wir dem beobachteten Verhalten der Shifts ein Zwei- Zustands- Modell zugrunde gelegt haben, das auf der Annahme beruht, dass eine Wasserstoffbrückenbindung bei zunehmender Temperaturerhöhung und damit verbundener struktureller Änderung eine „Dehnung“ erfährt, die mit Shiftänderungen verbunden ist. Beim Aufbrechen der H-O- Bindung ergeben sich stärkere Änderungen der chemischen Verschiebung

Schwierig erwies sich bei der dabei verwendeten Formel (vgl. Gleichung 9) die Definition des Parameters  $T_0$ , da nicht ohne weiteres erkenn- und begründbar war, welchen Wert dieser annehmen sollte. Einerseits könnte er den absoluten Nullpunkt bei 0K repräsentieren, andererseits den Gefrierpunkt bei 0°C, also 273K. Auch denkbar wäre es, dass in diesem Zusammenhang der Nullpunkt bezüglich der Temperatur den Punkt der Kältedenaturation des verwendeten Proteins angibt. Um den Einfluss zu untersuchen, den die Wahl von  $T_0$  auf die Fitfunktion ausübt, haben wir die Daten zunächst mit unterschiedlichen Werten von  $T_0$  angenähert. Dabei wurde untersucht, welche Kombinationen aus  $\Delta G_0$  und  $\Delta\delta$  bei einem

gegebenen  $T_0$  die besten Werte in der Summe der quadratischen Abweichungen aufweisen, was ein Maß dafür ist, wie gut die Daten durch den verwendeten Fit erklärt werden.

Anhand der durchgeführten Untersuchungen ließ sich nicht zweifelsfrei entscheiden, welchem Wert  $T_0$  entspricht, allerdings konnte mit einiger Sicherheit ausgeschlossen werden, dass in diesem Zusammenhang  $T_0$  den absoluten Nullpunkt bei 0K darstellt. Weitere Untersuchungen zur genaueren Bestimmung und zur generellen Überprüfung des verwendeten Modells sind geplant. Zu beachten ist dabei, dass der beobachtete Effekt klein im Vergleich zum vorhandenen Rauschen ist, was die exakte Quantifizierung erschwert.

Anzumerken ist weiterhin, dass die Shiftänderungen nicht allein von fehlenden oder gebildeten Wasserstoffbrückenbindungen hervorgerufen werden, sondern auch kleine lokale Veränderungen in der Struktur den Einfluss von OH- Bindungen überlagern.

Nicht zu vernachlässigen ist auch das Problem des Fits der Daten mit der im Text beschriebenen Formel, die für ein eindeutiges Ergebnis zu viele Freiheitsgrade aufweist, sodass Parameter vorgegeben werden müssen (z.B.  $T_0$  oder  $\Delta\delta$ ). Auch dann noch sind diese Berechnungen sehr rechenintensiv und zeitaufwendig.

Für detailliertere Aussagen in diesem Bereich müssen einige Parameter experimentell bestimmt (z.B. experimentelle Bestimmung des Kältedenaturationspunktes) und/oder eine Erweiterung des Daten- und Messbereiches angestrebt werden. Messungen bis hin zu Kälte- und Hitzedenaturation könnten Aufschluss darüber geben, wie sich die chemischen Verschiebungen an den Randbereichen verhalten, und welches mathematische Modell, bzw. welches  $T_0$  verwendet werden soll.

Diese Messungen sind allerdings momentan wegen technischer Gegebenheiten nicht möglich, da mit dem Vorhandenen Kryo- Probenkopf zwar sehr rauscharme und signalstarke Spektren aufgenommen werden können, allerdings wegen des herrschenden Temperaturunterschiedes zwischen Probe und Messkopf keine Temperaturen über 60 °C und unter 5°C vermessen werden können.

Nichtsdestoweniger haben die gezeigten Untersuchungen die Hoffnungen darin bestärkt, dass das Temperaturverhalten der chemischen Verschiebungen mathematisch vorhergesagt werden kann. Außerdem zeigten sich deutliche Abweichungen des Shiftverhaltens von der Linearität, was zwar seit längerem postuliert, aber hier erstmals an einer umfangreichen Messserie experimentell bestätigt wurde.

## 7. Zusammenfassung

Die vorliegende Arbeit besteht im Wesentlichen aus zwei Teilen. Im ersten Teil wurden die Programme STEREO und AUNOAS entwickelt, bzw. weitergeführt. STEREO ermöglicht dabei die Verbesserung der stereospezifischen Zuordnung von NMR-Signalen. Damit trägt es zu einer weiteren Steigerung der Genauigkeit von Proteinstrukturen, die mit Hilfe von NMR-Messungen ermittelt werden, bei. Mit dem Programmpaket AUNOAS wurde die Basis für den experimentellen Teil der Arbeit gelegt. Dieses Programm gestattete es erstmals, die sehr zeitaufwendige Arbeit des manuellen Zuordnens von NMR- NOESY Spektren computerisiert durchzuführen. Damit wurde es möglich, NMR-Strukturen in wesentlich kürzerer Zeit zu ermitteln. Dieses Potential des Programms AUNOAS wurde anschließend im experimentellen Teil der Arbeit verwendet, um thermisch angeregte Zustände von Proteinen zu untersuchen. Dazu wurden hochaufgelöste NMR- NOESY Spektren des Proteins Saratin bei verschiedenen Temperaturen aufgenommen und die dazugehörigen Proteinstrukturen ermittelt. Es zeigte sich, dass mit den dabei gewonnenen Daten diverse theoretische Vorstellungen experimentell bestätigt werden konnten.

Wie bereits erwähnt, wurden in dieser Promotion unter anderem die Arbeiten an dem Programm STEREO weitergeführt und zum Abschluss gebracht. STEREO dient zur Strukturverbesserung, namentlich zur stereospezifischen Zuordnung von Methylen- und Methylprotonen. Das Programm vollzieht die stereospezifische Zuordnung aufbauend auf dem bereits beschriebenen Programm RFAC [Gro00]. Es ermittelt diejenigen stereospezifischen Zuordnungen, die durch Vergleich von rückgerechneten und experimentellen Spektren, und daraus resultierenden Übereinstimmungen oder Diskrepanzen eine Fehlzusammenordnung aufweisen [Kir01].

Zur besseren Handhabbarkeit und zur Beschleunigung des Programmablaufes wurde die Routine zur Berechnung der R- Werte und zum Vergleich von experimentellen und simulierten Spektren in den Programmablauf integriert, und aufgrund geänderter Eingabedatenformate nicht mehr benötigte Routinen entfernt. Außerdem wurde die interne Datenbasis des Programms STEREO auf die in AUREMOL verfügbare globale Datenbank umgestellt. Dabei ist es jetzt auch möglich, mittels eines allgemein üblichen Codeschlüssels, z.B. stereospezifische Paare anzugeben, die manuell gesichert sind, und nicht verändert werden dürfen. Das Programm arbeitet jedoch weiterhin sowohl mit Daten im IUPAC, als auch im DYANA/CNS- Format.

Dieses Programm wurde erfolgreich an diversen Proteinen getestet, zusätzlich kam ein Bündel aus dreizehn experimentellen NMR- Strukturen zum Einsatz. Dabei zeigte sich deutlich, dass es vor allem bei Verwendung von mehreren Strukturen des gleichen Proteins möglich ist, statistisch gesicherte Aussagen darüber zu treffen, welche stereospezifisch zugeordneten Signale falsch benannt wurden, und welche nicht.

Bei der Betrachtung der verschiedenen R- Werte für Datensätze vor, und nach dem von STEREO durchgeführten Tausch, zeigten sich Verringerungen der R- Werte, die auf eine Verbesserung der Struktur hinweisen.

Die in diesem Zusammenhang durchgeführten statistischen Untersuchungen an Einträgen der *Protein Data Bank* über die chemischen Verschiebungen von stereospezifisch betrachteten Signalen einzelner Aminosäuren brachten interessante Ergebnisse. Wie sich zeigte, besteht ein deutlicher Zusammenhang zwischen kleinerer chemischer Verschiebung und der Namensgebung als „kleinerer Name“ z.B.  $H^{\beta 3}$ . Dieses Resultat bedarf allerdings weitergehender Untersuchungen, um auszuschließen, dass es sich um menschliche Fehler bei der Benennung, oder Bereitstellung der Daten in den Datenbanken handelt.

Das im zweiten Teil der Arbeit entwickelte Programm zur automatischen Zuordnung von NOESY- Spektren, AUNOAS, hat sich als sehr stabiles und hilfreiches Tool zur Bewältigung der zeitraubenden Arbeit des Benennens von Spinsystemen und Spektren erwiesen. Der iterative, auf sich selbst rückkoppelnde Mechanismus, der weitestgehend die menschliche Vorgehensweise beim Benennen von NOESY- Spektren nachvollzieht, ermöglicht eine außergewöhnlich fehlerfreie automatische Zuordnung der Spektren, da der Algorithmus bei Unsicherheiten in der Zuordnung primär mit der Nichtzuordnung von Signalen reagiert, nicht aber mit falscher Benennung. Diese Tatsache ist nicht zu unterschätzen, da ein falsch zugeordnetes Signal viel schwerer wiegt als etliche unzugeordnete.

Wie bei diversen Tests zur Stabilität und Fehleranfälligkeit nachgewiesen wurde, zeigte das Programm eine sehr hohe Zuordnungswahrscheinlichkeit bei einer nahezu vernachlässigbaren Fehlerquote. Selbst vergleichsweise „riesige“ Fangradienten von 0,2 ppm bei einem Spektrum mit 6000 Peaks führen zu einem guten Ergebnis. Ebenfalls bei verrauschten Masterlisten (zufällig leicht verschobene Position der einzelnen Signale) ordnete das Programm unter realistischen Bedingungen nahezu 100% der Peaks richtig zu.

Auch das Vorhandensein von Artefakten, Rauschsignalen oder Verunreinigungen in den Spektren beeinträchtigte den Programmablauf nicht. Egal, um welchen Störungstyp es sich handelte, und auch weitgehend unabhängig von der absoluten Anzahl der vorhandenen Störungen. In allen Fällen wurden die Signale des verwendeten Proteins richtig und vollständig zugeordnet.

Diese Tatsache ermöglichte es, hochaufgelöste experimentelle NOESY- Spektren mit hohem Signalaufkommen vollständig automatisch zuzuordnen, und dadurch die angestrebte Beobachtung thermisch angeregter Zustände eines Proteins durchzuführen.

Diese Beobachtungen wurden im letzten Teil der Arbeit, auch als Anwendungsbeispiel für die Funktionsweise des Programms AUNOAS dargestellt. Dazu wurde eine Temperaturreihe des Proteins Satatin vermessen, 41 Spektren aufgenommen, prozessiert, mit AUNOAS zugeordnet und räumliche Strukturen für jede gemessene Temperatur berechnet.

Aufgrund der automatischen Zuordnung der Spektren war es möglich, diese Arbeit in etwa drei Monaten zu bewerkstelligen, die unter herkömmlichen Bedingungen mittels manueller Zuordnung einige Jahre dauern würde.

Die erhaltenen Daten ermöglichten erstmals Einblick in einen weiten Bereich thermischer Strukturdynamik. Es zeigte sich, dass das verwendete Protein, Saratin, über den gesamten beobachteten Bereich wohlgefaltet ist. Denaturierungserscheinungen traten nicht auf. Dennoch konnten strukturelle Änderungen in Abhängigkeit von der thermischen Anregung über einen weiten Bereich mit Hilfe der NMR-Spektroskopie sichtbar gemacht werden. Damit konnten schon länger diskutierte theoretische Betrachtungen zum Verhalten von Proteinen in der Nähe ihres Faltungsenergieminimums- „am Grund des Faltungstrichters“- erstmals experimentell beobachtet werden.

Auch das Verhalten der chemischen Verschiebungen der  $^1\text{H}$ - Signale nahezu aller vorhandenen Atome bei zunehmender Temperatur wurde untersucht, und ein mathematisches Modell zu ihrem Verlauf vorgeschlagen. Es zeigte sich, dass der Verlauf vieler Shiftänderungen, wie seit längerem postuliert, nicht exakt linear ist. In dieser umfangreichen Messserie konnte dies anhand sehr präziser Daten gezeigt werden.

Im Rahmen dieser Arbeit konnte gezeigt werden, dass hochaufgelöste NMR- Spektroskopie zusammen mit speziell auf die Probleme der NMR zugeschnittenen Software- Algorithmen in der Lage ist, dynamische Prozesse von Proteinen wissenschaftlich zugänglich zu machen. Dies kann in Zukunft zusammen mit der stetig voranschreitenden Entwicklung von Geräten die NMR als Methode zur Strukturbestimmung wieder in den Vordergrund rücken. Insbesondere die Möglichkeit auch dynamische Prozesse in ihrer natürlichen Umgebung zu analysieren dürfte sich hier als großer Vorteil erweisen.

## 8. Literatur

- [ And97 ] Andersen N. H., Neidigh J. W., Harris S. M., Lee G. M., Liu Z. H., Tong H. (1997). J. Am. Chem. Soc. 119, 8547- 8561.
- [ Ant95 ] Antz C., Neidig K.P., Kalbitzer H.R. (1995). J. Biomol. NMR. 5, 287- 296.
- [ Alb97 ] Alberts B., Bray D., Lewis J., Raff M., Roberts K., Watson J.D. (1997). "Molekularbiologie der Zelle", VCH, Weinheim.
- [ Asa95 ] Asakura T., Taoka K., Demura M., Williamson M. P. (1995). J. Biomol. NMR 6, 227- 236.
- [ Bam01 ] Bamann C. (2001). Diplomarbeit. Universität Regensburg.
- [ Bax97 ] Baxter N. J., Williamson M. P. (1997). J. Biomol. NMR 9, 359- 369.
- [ Bax98 ] Baxter N. J., Hosszu L. L. P., Waltho J. P., Williamson M. P. (1998). J. Mol. Biol. 284, 1625- 1639.
- [ Ben94 ] Beneicke W. (1994). Dissertation. Universität Heidelberg.
- [ Bio04 ] [http://www.bmrb.wisc.edu/elec\\_dep/gen\\_aa.html](http://www.bmrb.wisc.edu/elec_dep/gen_aa.html)
- [ Bmr01 ] BioMagRes data bank (2001). <http://www.bmrb.wisc.edu/>
- [ Bmr04 ] BioMagRes data bank (2004). <http://www.bmrb.wisc.edu/>
- [ Bru87 ] Brünger A.T., Campbell R.L., Clore G.M., Groneborn A.M., Karplus M., Petsko G.A., Teeter M.M. (1987). Science. 235, 1049- 1053.
- [ Bru92 ] Brünger A.T. (1992). X-PLOR Manual Version 3.0, Yale University Press, New Haven.
- [ Bru98 ] Brünger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.-S., Kuszewski J., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T, und Warren G.L. (1998) Acta Cryst., D54, 905-924.
- [ Bru04 ] <http://www.bruker.de>
- [ Cav95 ] Cavanagh J., Fairbrother W. J., Palmer III A.G., Skelton N. J. (1995). "Protein NMR Spectroscopy: Principles and Practice", Academic Press.
- [ Cie01 ] Cierpicki T., Otlewski J. (2001). J. Biomol. NMR 21, 249- 261.



- [ Cor67 ] Cornfield J. (1967) Rev. Int. Statist. Inst., 35, 34-49.
- [ Cor69 ] Cornfield J. (1969) Biometrics, 25, 617-642.
- [ Cor99 ] Cornilescu G., Delagio F. and Bax A. (1999). J. Biomol. NMR, 13, 289-302.
- [ Cor02 ] Cordier F., Grzesiek S. (2002). 317, 739- 752.
- [ Dom04 ] Domogalla B. (2004). Dissertation. Universität Regensburg, in verb.
- [ Els04 ] Elsner R. (2004). Dissertation. Universität Regensburg, in verb.
- [ Ern66 ] Ernst R.R. (1966). Adv. Magn. Reson. 2, 1- 135.
- [ Ern87 ] Ernst R. R., Bodenhausen G. and Wokaun A. (1987). "Principles of Nuclear Magnetic Resonance in One and Two Dimensions", Clarendon Press, Oxford.
- [ Eva95 ] Evans J. N. S. (1995). "Biomolecular NMR Spectroscopy", Oxford University Press.
- [ Fer78 ] Ferrige, A. G. and Lindon, J. C. (1978). J. Magn. Reson., 31, 337-340.
- [ Fra87 ] Frauenfelder H., Hartmann H., Karplus M., Kuntz I. D., Kuriyan J., Parak F., Petsko G. A. Ringe D., Tilton R. F., Conolly M. L., Max N. (1987). Biochem. 26, 254- 261.
- [ Fri92 ] Friebolin H., (1992). "Ein- und zweidimensionale NMR-Spektroskopie. Eine Einführung", VCH-Verlag.
- [ Fol97 ] Folmer R. H. A., Hilbers C. W., Konnings R. N. H., Nilges M. (1997). J. Biomol. NMR, 9, 245-258.
- [ Gan02 ] Ganslmeier B. (2002). Dissertation. Universität Regensburg.
- [ Gey95 ] Geyer M., Neidig K.-P., Kalbitzer H.R. (1995) J. Magn. Reson. Ser. B, 109, 31-38.
- [ Gey97 ] Geyer M., Herrmann C., Wohlgemuth S., Wittinghofer A. und Kalbitzer H.R. (1997). Nat. Struct. Biol. 4, 694-699.
- [ Goe94 ] Görler A. (1994). Diplomarbeit. MPI für medizinische Forschung. Heidelberg.
- [ Goe97 ] Görler A., und Kalbitzer H.R. (1997). J. Magn. Reson., 124, 177-188.
- [ Goe99a ] Görler A., Gronwald W., Neidig K. P., Kalbitzer H. R. (1999). J. Magn. Reson. 137, 39-45.



- 
- [ Goe99b ] Görler A., Hengstenberg W., Kravanja M., Beneicke W., und Kalbitzer H. R. (1999). Appl. Magn. Reson., 17, 465-480.
- [ Gri87 ] Griesinger C., Sorensen O. W., Ernst R. R. (1987). J. Magn. Reson. 73, 574- 579.
- [ Gro00 ] Gronwald W., Kirchhöfer R., Görler A., Kremer W., Ganslmeier B., Neidig K. P., und Kalbitzer H. R. (2000). J. Biomol. NMR, 17, 137-151.
- [ Gro01 ] Gronwald W., Brunner E., Huber F., Wenzler M., Hermann C., Kalbitzer H. R. (2001). Protein Science 10, 1260- 1263.
- [ Gro02 ] Gronwald W., Moussa S., Elsner R., Jung A., Ganslmeier B, Trenner J., Kremer W., Neidig K. P., Kalbitzer H. R. (2002). Journal of Biomol. NMR, 23, 271- 278.
- [ Gro04 ] Gronwald W., Brunner K., Kirchhöfer R., Nasser A., Trenner J., Kalbitzer H. R., Ganslmeier B., Riepl H., Scheiber J., Görler A., Ried A., Elsner R., Neidig K. P. (2004). Manual AUREMOL, Universität Regensburg.
- [ Gue89 ] Güntert P., Braun W., Billeter M. und Wüthrich K. (1989). J. Am. Chem. Soc., 111, 241-263.
- [ Gue91 ] Güntert P., Quian P. Q., Otting G., Müller M., Gehring W. und Wüthrich K. (1991). J. Mol. Biol., 217, 531-540.
- [ Gue97 ] Güntert P., Mumenthaler C., Wüthrich K. (1997) J. Mol. Biol. , 273, 283-298.
- [ Har98 ] Harrieder S. (1998). Diplomarbeit. Universität Regensburg.
- [ Hau89 ] Hausser H., Kalbitzer H. R. (1989). "NMR in Medicine and Biology", Springer Verlag, Berlin Heidelberg.
- [ Hom89 ] Homan S. (1989). "A Dictionary of Concepts in NMR", Oxford Science Publications.
- [ Hel00 ] Helgstrand M., Kraulis P., Allard P., Härd T. (2000). Journal of Biomol. NMR, 18, 329-336.
- [ Her02a ] Herrmann T., Güntert P., Wüthrich K. (2002). J. Mol. Biol. 319, 209- 227.
- [ Her02b ] Herrmann T., Güntert P., Wüthrich K. (2002). Journal of Biomol. MNR, 24, 171- 189.
- [ Hit03 ] Hitchens T. K., Lukin J. A., Zhan Y., McCallum S. A., Rule G. S. (2003). Journal of Biomol. NMR, 25, 1- 9.
- [ Hub01 ] Huber F. (2001). Dissertation. Universität Regensburg.
- [ Jee71 ] Jeener J. (1971). II. Ampere International Summer School, Basko Polje, Jugoslawien.
- [ Jee79 ] Jeener J., Meier B. H., Bachmann P. und Ernst R. R. (1979) J. Chem. Phys. 71, 4546-4553.
-

- [ Kal04 ] Kalbitzer H.R. (2004). Persönl. Mitteilung.
- [ Kap85 ] Kaptein R., Zuiderweg E. R. P., Scheek R. M., Boelens R., van Gunsteren W. F. (1985). J. Mol Biol. 182, 179- 182.
- [ Kay90 ] Kay L. E., Clore G. M., Bax A., Groneborn A. M. (1990). Science 249, 411- 414.
- [ Kes88 ] Kessler H., Gehrke M., Griesinger C. (1988). Angewandte Chemie. 100, 507.
- [ Kee84 ] Keepers J. W., James T. L. (1984). J. Magn. Reson. 57, 404- 426.
- [ Kha99 ] Khare D., Alexander P., Orban J. (1999). Biochem. 38, 3918- 3925.
- [ Kir01 ] Kirchhöfer R. (2001). Diplomarbeit. Universität Regensburg.
- [ Ker90 ] Kernighan B., Ritchie D. (1990). "Programmieren in C", Hanser Verlag München.
- [ Kir83 ] Kirkpatrick S., Gelatt, C. D., Vecchi M. P. (1983). Science 220, 671- 680.
- [ Kre01 ] Kremer W., Schuler B., Harrieder S., Geyer M., Gronwald W., Welker C., Jaenicke R., und Kalbitzer H. R. (2001). Eur. J. Biochem., 268, 2527-2539.
- [ Kus96 ] Kuszewski J., Groneborn A. M., Clore G. M. (1996). J. Magn. Res. Ser. B, 112, 79- 81.
- [ Las98 ] Laskowski R.A., Mac Arthur M.W., Thornthorn J.M. (1998). Curr. Opin. Struct. Biol. 8, 631- 639.
- [ Loe99 ] Löffler G. (1999). "Basiswissen Biochemie mit Pathobiochemie", Springer, Berlin.
- [ Lip82 ] Lipari G., Szabo A. (1982). J. Am. Chem. Soc. 104, 4546- 4559.
- [ Mac80 ] Macura S. and Ernst R. R. (1980). Mol. Phys. 41(1), 95.
- [ Mar83 ] Marion D. und Wüthrich K. (1983). Biochem. Biophys. Res. Comm. 113, 967-974.
- [ Mau00 ] Maurer T. (2000). Nachrichten aus der Chemie 48, 1336- 1341.
- [ Mol04 ] MOLMOL 2K.1, Institut für Molekularbiologie und Biophysik, ETH Zürich, Spectrospin AG, Faellanden, Schweiz.
- [ Mor92 ] Morris A. L., MacArthur M. W., Hutchinson E. G., Thornthorn J. M. (1992). Prot. Struct. Funct. Gen., 12, 345- 364.

- 
- [ Mos95 ] Moseley H. N. B., Curto E. V. und Krishna N. R. (1995). J. Magn. Reson. B 108, 243.
- [ Mye96 ] Myers J. K., Pace C. N. (1996). Biophys. J. 71, 2033- 2039.
- [ Nei90 ] Neidig K.P., Saffrich R., Lorenz M., Kalbitzer H.R. (1990). J. Magn. Reson. 89, 543- 552.
- [ Nei95 ] Neidig K.P., Geyer M., Görler A., Antz A., Saffrich A., Beneicke W., Kalbitzer H.R. (1995). J. Biomol. NMR. 6, 255- 270.
- [ Neu89 ] Neuhaus D., Williamson M. (1989). "The Nuclear Overhauser Effect in Structural and Conformational Analysis", VCH, New York, Weinheim, Cambridge.
- [ Nil91 ] Nilges M., Habazettl J., Brünger A.T., Holak T.A. (1991). J. Mol. Biol. 219, 499- 510.
- [ Nil93 ] Nilges M. (1993). Proteins. 17, 297.
- [ Nil90 ] Nilges M., Clore G. M. and Gronenborn A. M. (1990). Biopol. 29, 813-822.
- [ Oez02 ] Oezguen N., Adamian L., Xu Y., Rajarathnam K., Braun W. (2002). Journal of Biomol. NMR, 22, 249- 263.
- [ Pal97 ] Palmer III A. G. (1997). Current Opinion in Struct. Biol. 7, 732- 737.
- [ Per97 ] Pervushin K., Riek R., Wider G., Wüthrich K. (1997). Proc. Natl. Acad. Sci. USA 94, 12366- 12371.
- [ Pow77 ] Powell M. J. D. (1977). Mathematical Programming 12, 241.
- [ Pol95 ] Polshakov V. I., Frenkiel T. A., Birdsall B., Soteriou A. and Feeney, J. (1995). J. Magn. Reson. B 108, 31-43.
- [ Pri01 ] Pristovesek P., Rüterjans H., Jerala R. (2001). Journ. Computational Chem., 23, 335- 340.
- [ Ref03 ] Refaee M., Tezuka T., Akasaka K., Williamson M. P. (2003). J. Mol. Biol. 327, 857- 865.
- [ Rho99 ] <http://www.usm.maine.edu/~rhodes/CMCC/index.html>
- [ San82 ] Sandstrom J. (1982). "Dynamic NMR Spectroscopy", Academic Press, London.
- [ Sch97 ] Schulte A.C., Görler A., Antz C., Neidig K.P., Kalbitzer H.R. (1997) J. Magn. Reson., 129, 165-172.
- [ Sea92 ] Searle M. S., Williams D. H., Gerhard U. (1992). J. Am. Chem. Soc. 114, 10697- 10704.

- [ Sol55 ] Solomon I. (1955). Phys. Rev. 99, 2.
- [ Ste01 ] Steiner G. (2001). Dissertation. Universität Regensburg.
- [ Tej99 ] Tejero R., Monleon D., Celda B., Powers R. und Montelione G. T. (1999). J. Biomol. NMR 15, 251-264.
- [ Til92 ] Tilton R. F., Dewan J. C., Petsko G. A. (1992). Biochem. 31, 2469- 2481.
- [ Tre04 ] Trenner J. (2004). Dissertation. Universität Regensburg, in vorh.
- [ Wag83 ] Wagner G., Pardi A., Wüthrich K. (1983). J. Am. Chem. Soc. 105, 5948- 5949.
- [ Wel99 ] Welker C., Böhm G., Schurig H. und Jaenicke R. (1999). Protein Sci., 8, 394-403.
- [ Wil03 ] Williamson M. P.(2003). Proteins, 53, 731- 739.
- [ Wil92 ] Williamson M. P., Asakura T. (1993). J. Magn. Reson., Series B 101, 63- 71.
- [ Wut83 ] Wüthrich K., Billeter M. und Braun W. (1983). J. Mol. Biol., 196, 949-961.
- [ Wut86 ] Wüthrich K. (1986). "NMR of Proteins and Nucleid Acids", John Wiley & Sons, New York.
- [ Yip91 ] Yip P. F. und Case D.A. (1991). "Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy", Plenum Press New York.

## 9. Anhang

### 9.1. Vergleich der mittleren Strukturen

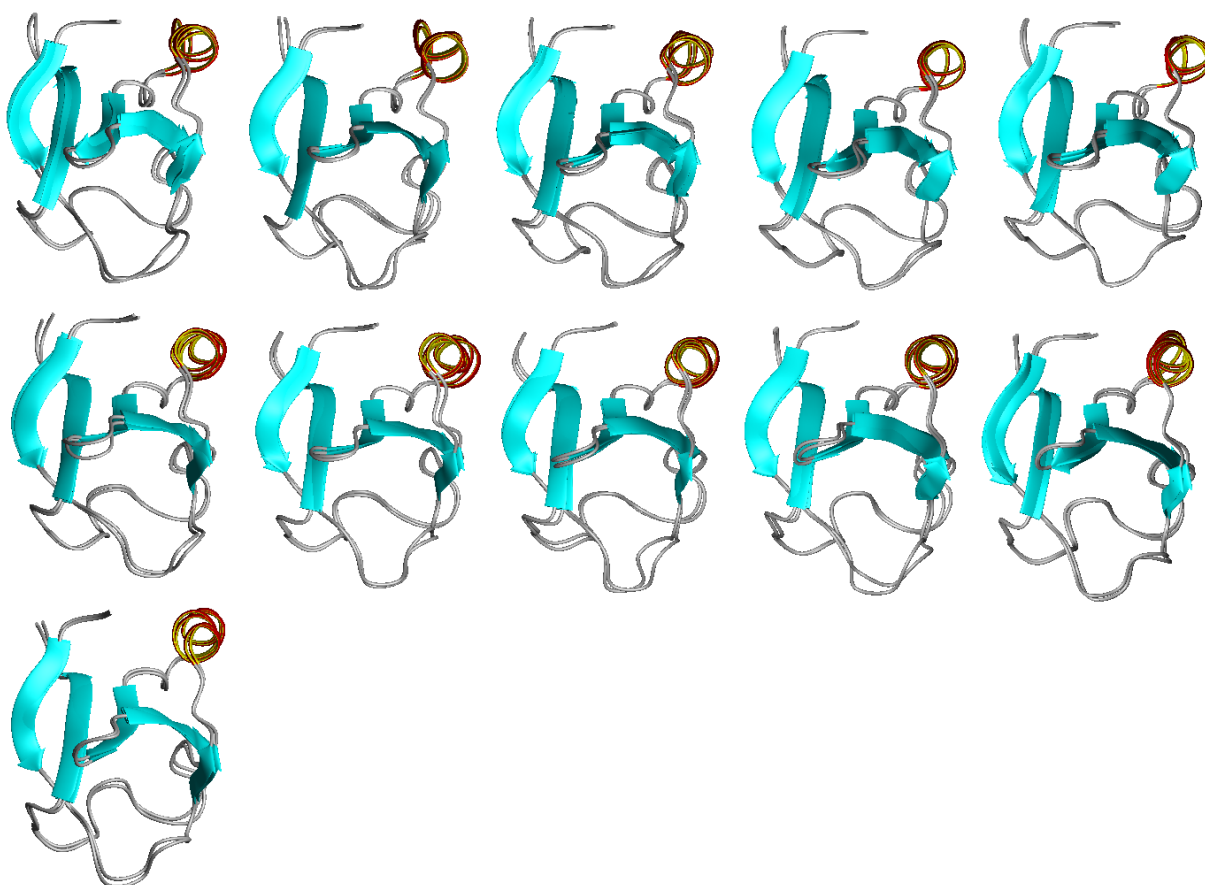
Vergleich der mittleren Struktur (aus dem Bündel der 10 Besten aus 1000 gerechneten Strukturen ) jeder Temperatur mit engen und weiten Fehlergrenzen.

#### ***283K bis 292K***



#### ***293K bis 302K***



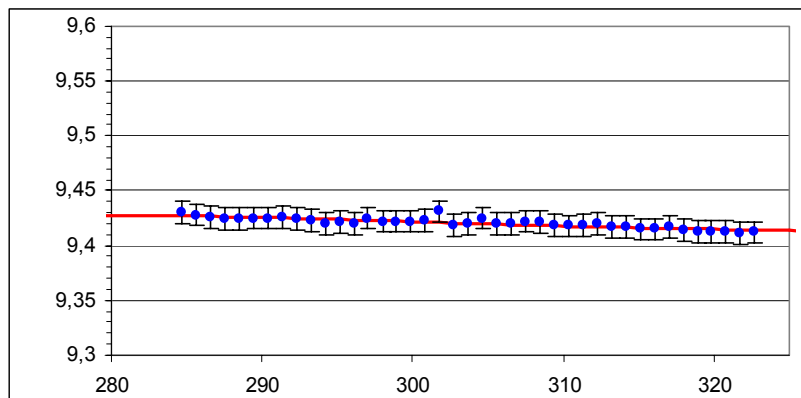
**303K bis 312K****313K bis 323K**

## 9.2. Verlauf der $H^N$ - Shifts über den Temperaturbereich

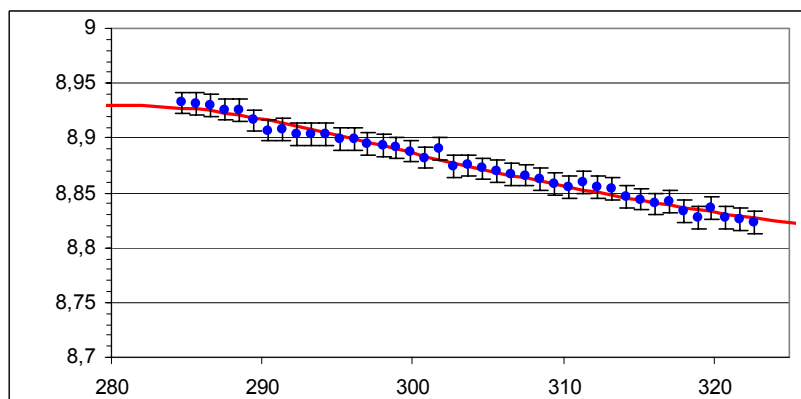
Alle  $H^N$ - Shifts über den Temperaturbereich verfolgt, mit zwei- Zustandsmodell gefittet.

Blau: Faltblatt, rot: Helix, farblos: Loop.

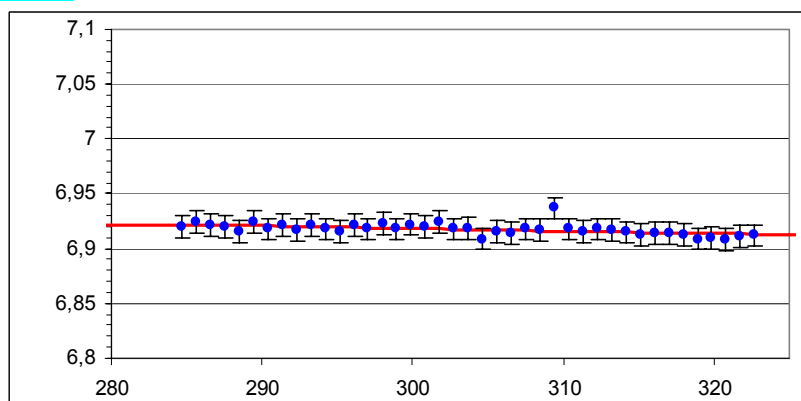
$H^N$  8

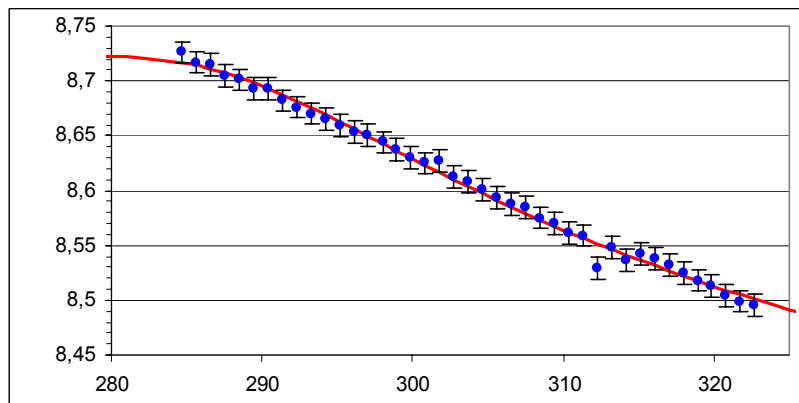
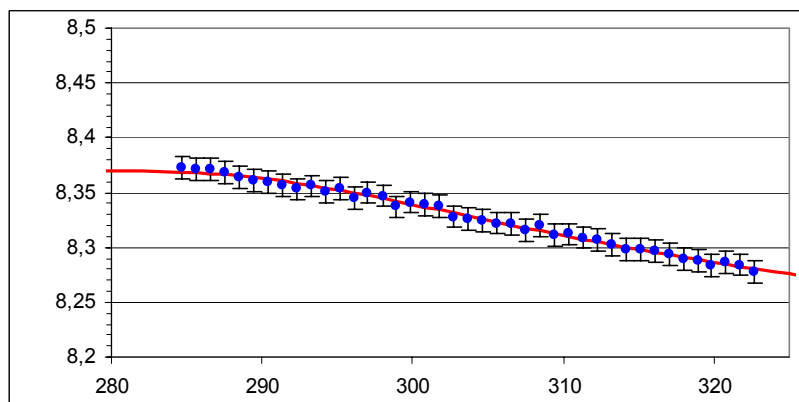
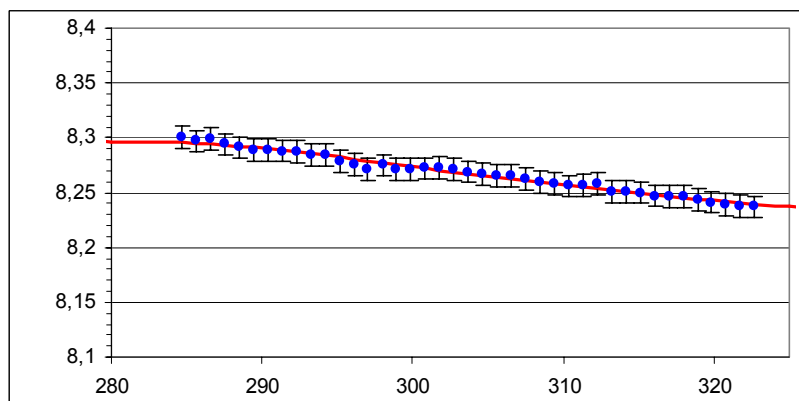


$H^N$  10

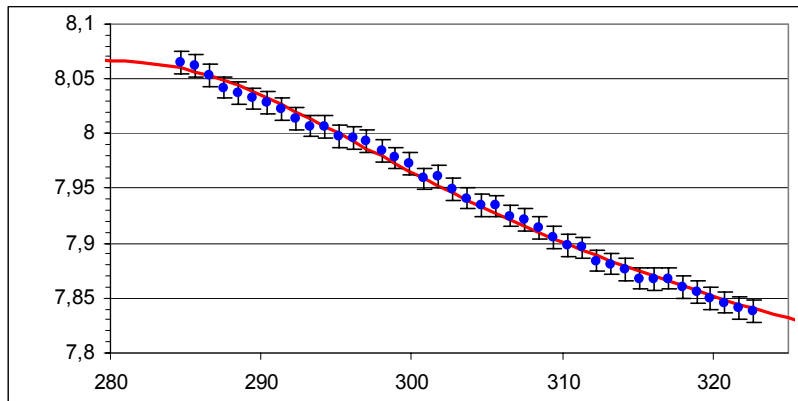
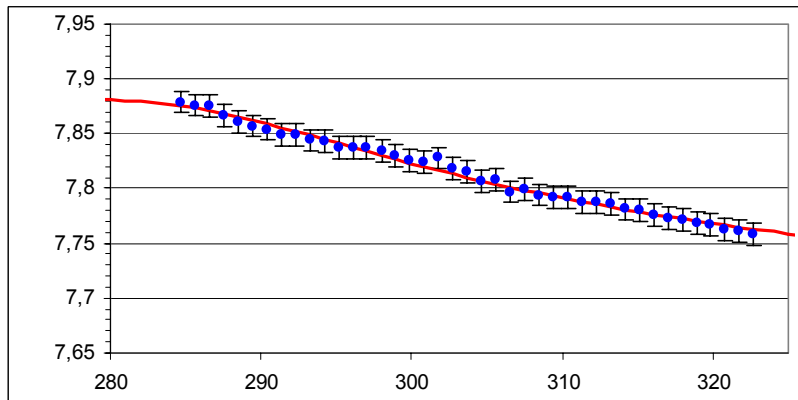
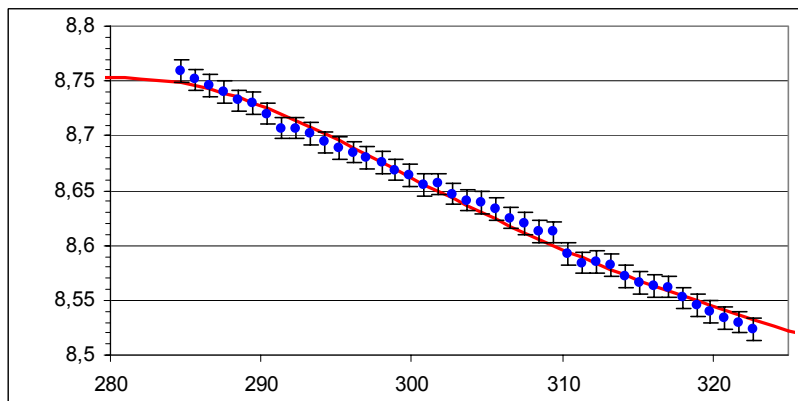


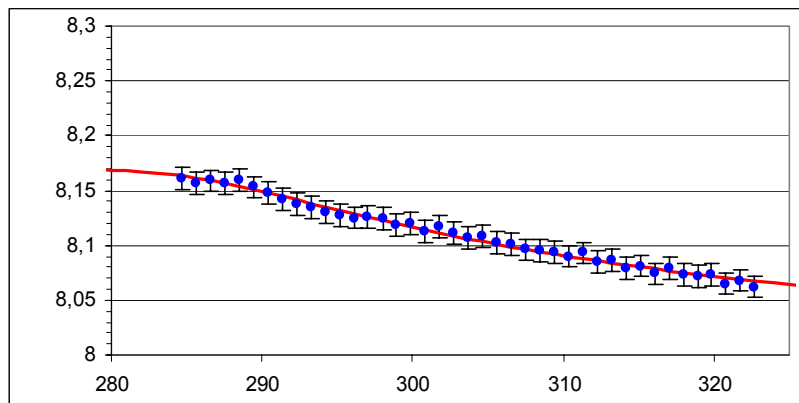
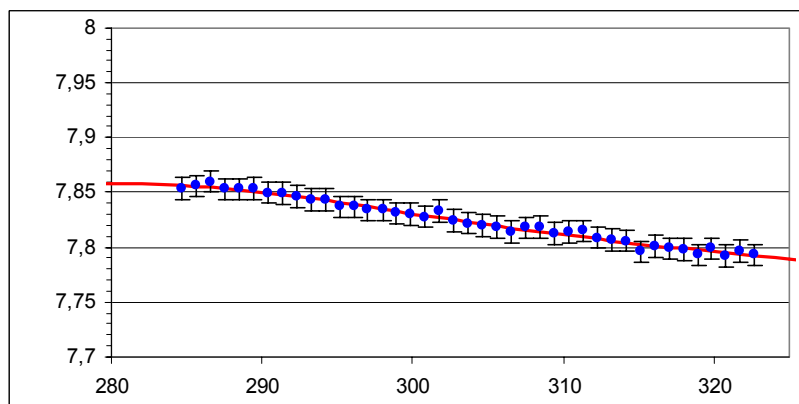
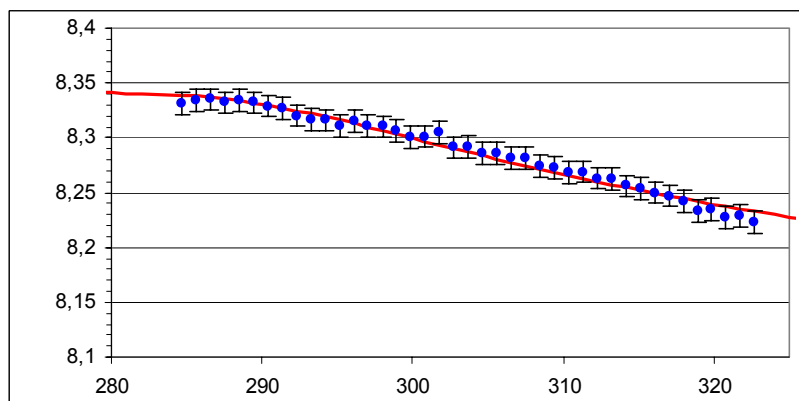
$H^N$  11

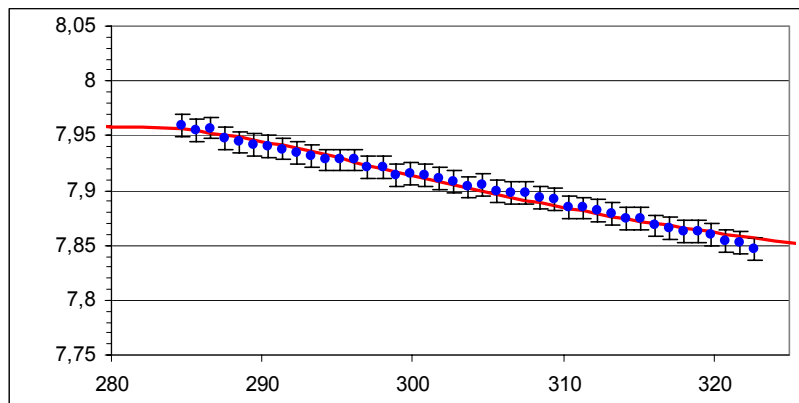
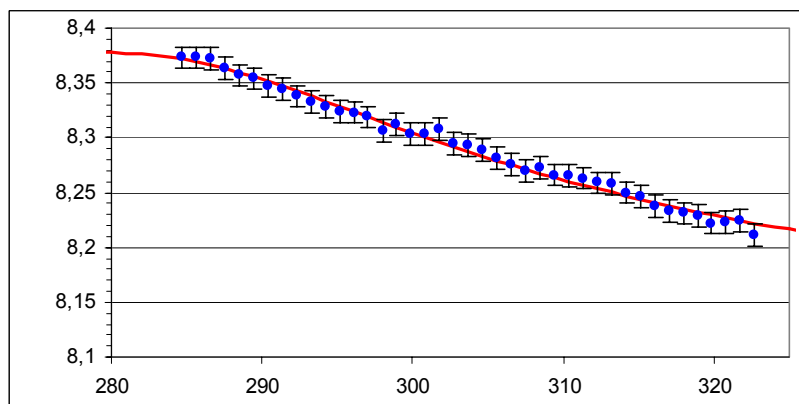
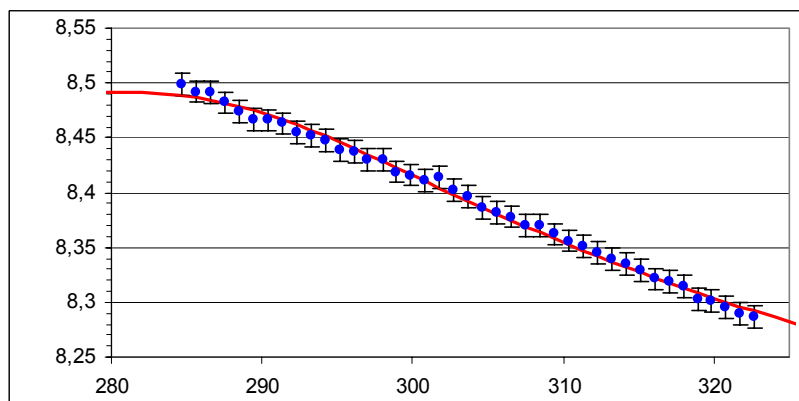


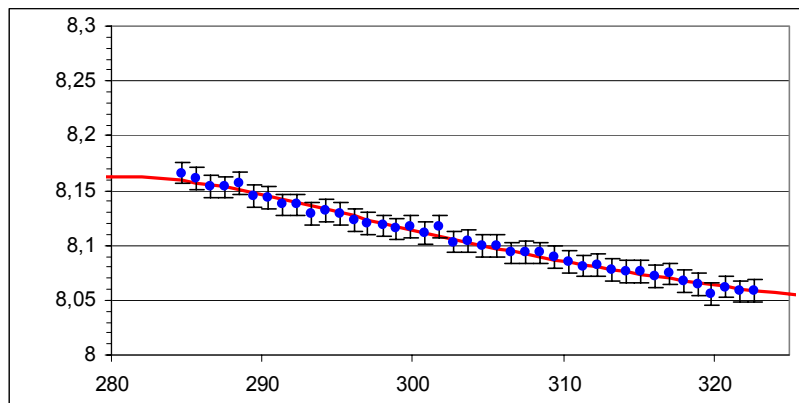
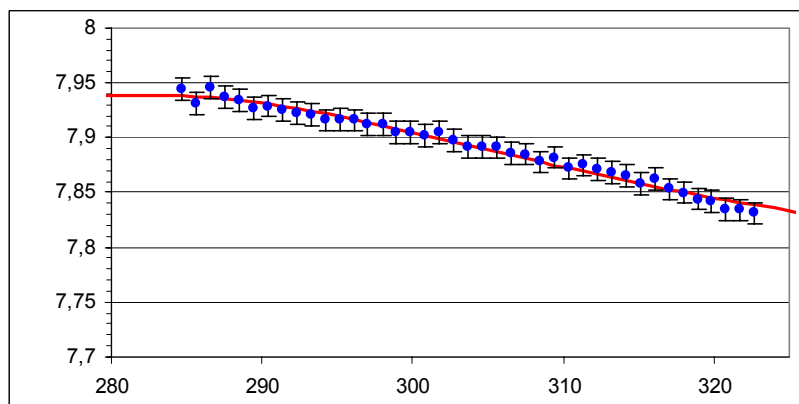
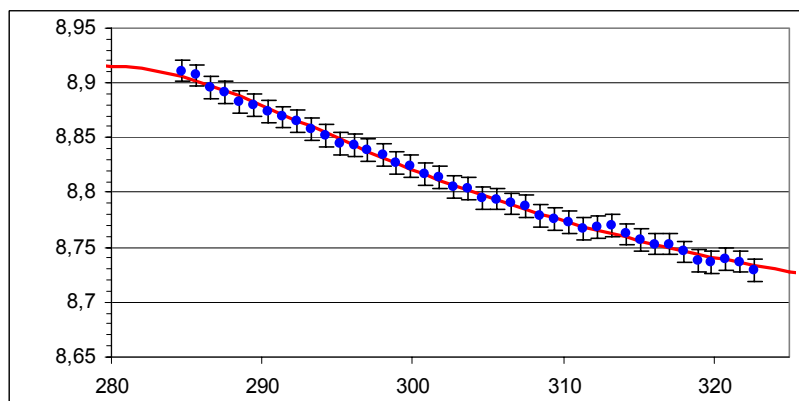
**H<sup>N</sup> 12****H<sup>N</sup> 22****H<sup>N</sup> 23**

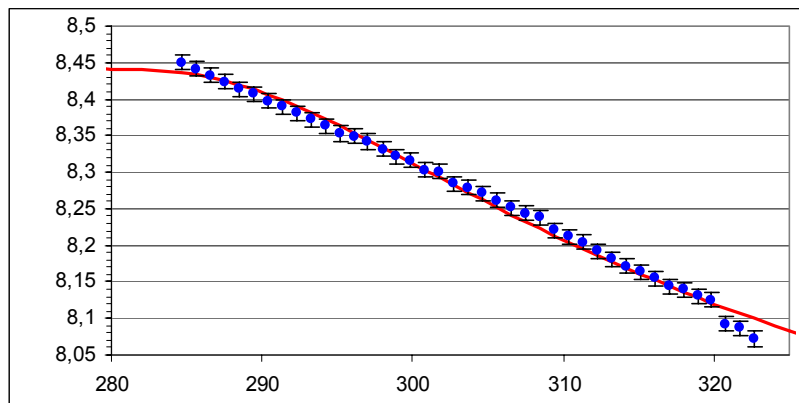
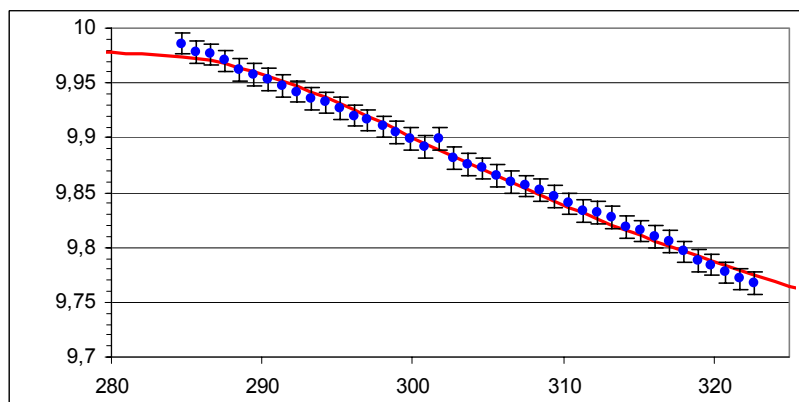
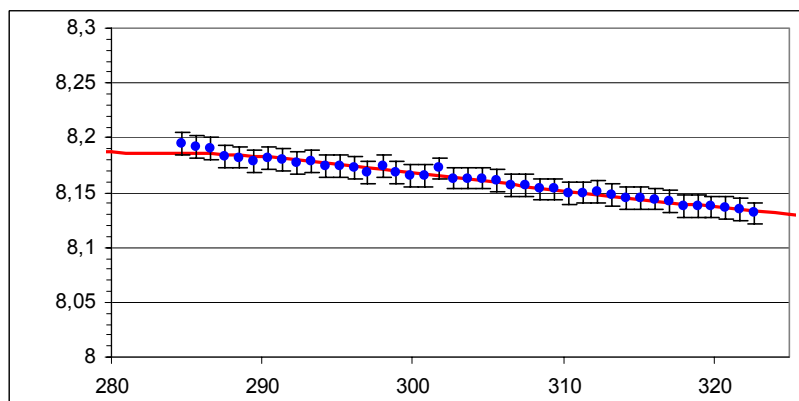


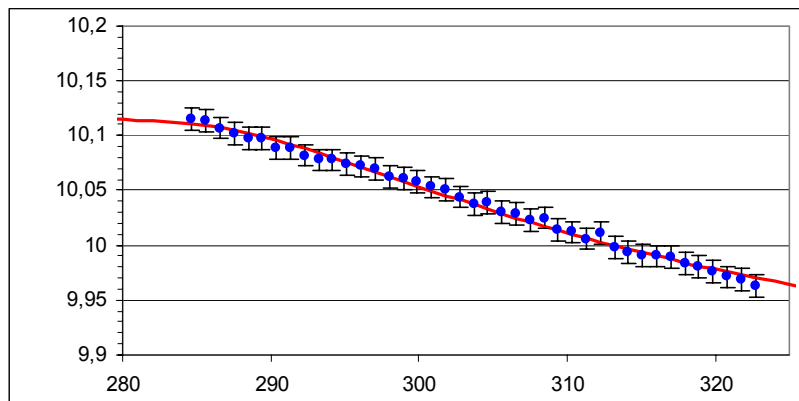
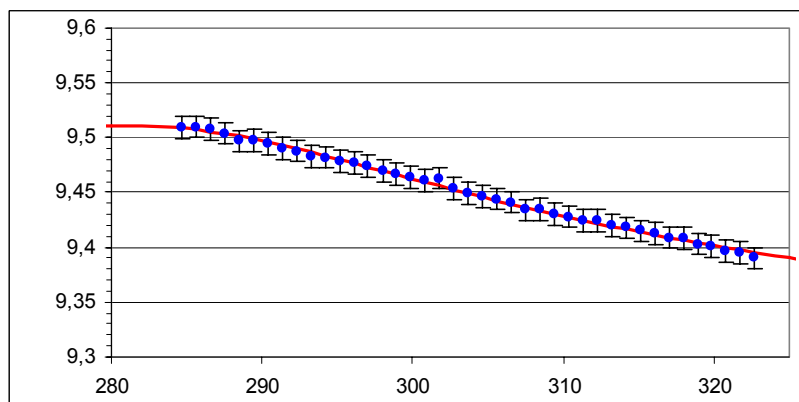
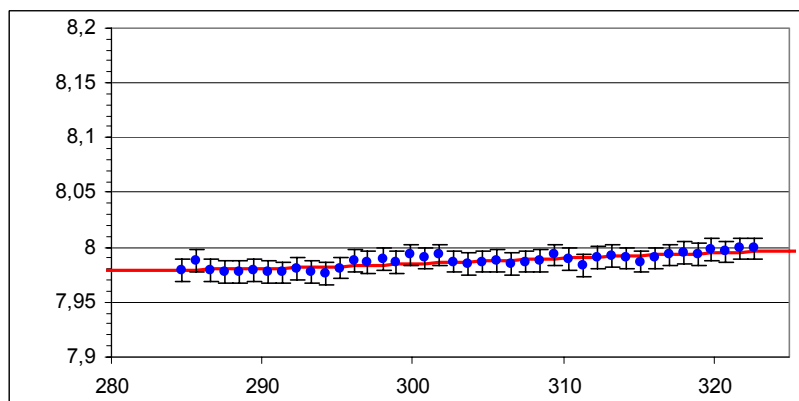
**H<sup>N</sup> 24****H<sup>N</sup> 27****H<sup>N</sup> 28**

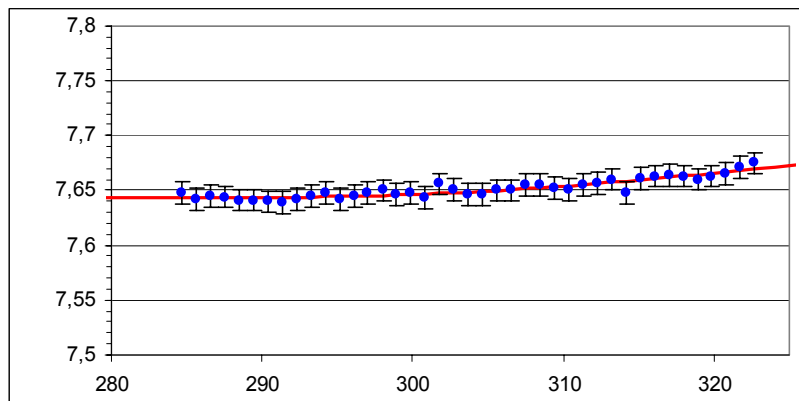
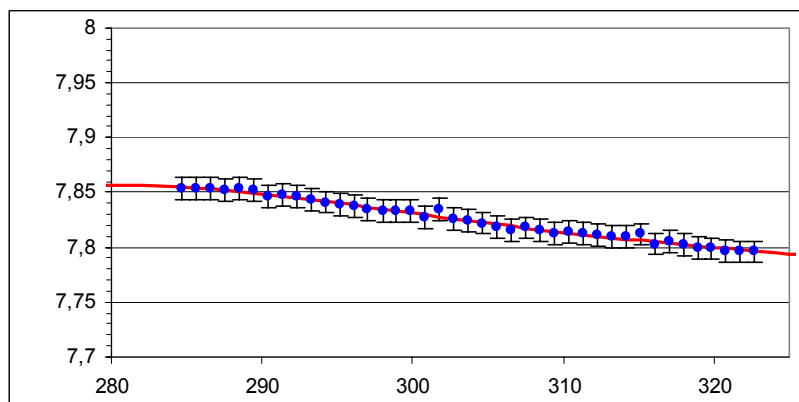
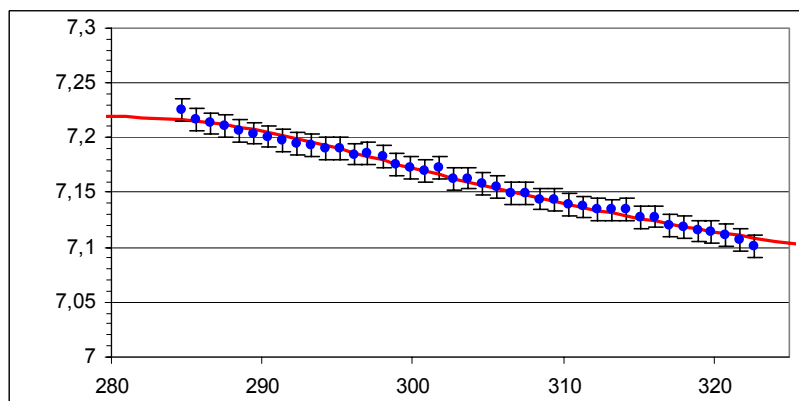
**H<sup>N</sup> 29****H<sup>N</sup> 30****H<sup>N</sup> 31**

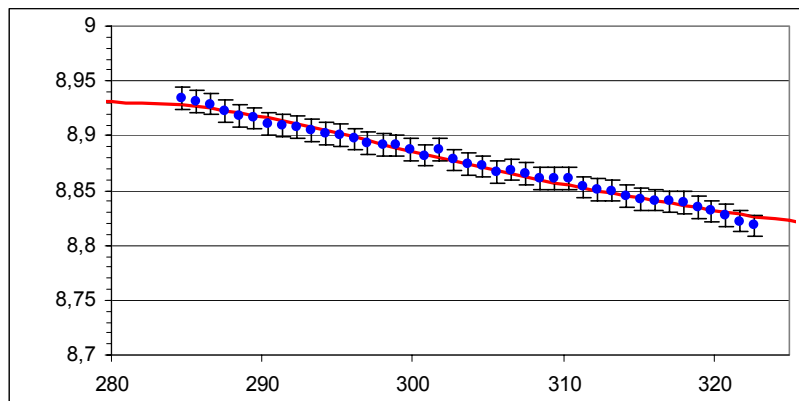
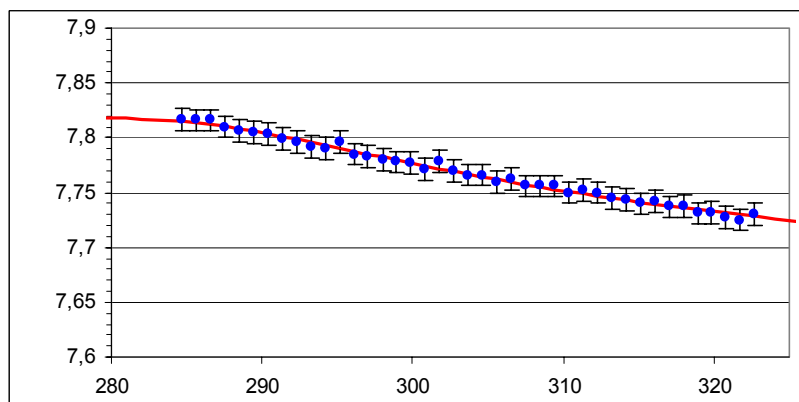
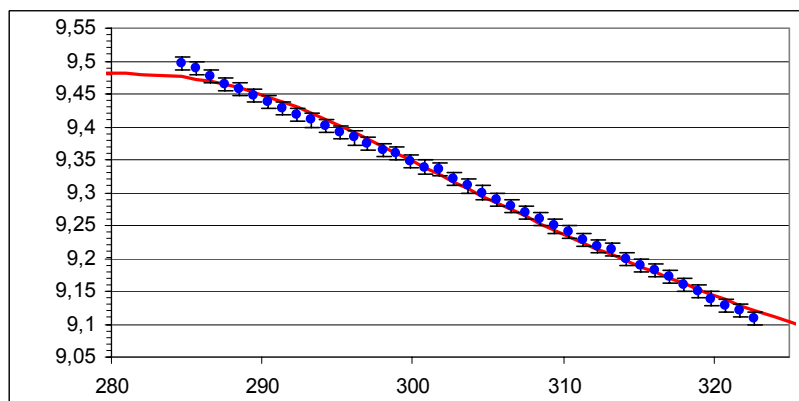
**H<sup>N</sup> 32****H<sup>N</sup> 34****H<sup>N</sup> 35**

$H^N 36$  $H^N 38$  $H^N 39$ 

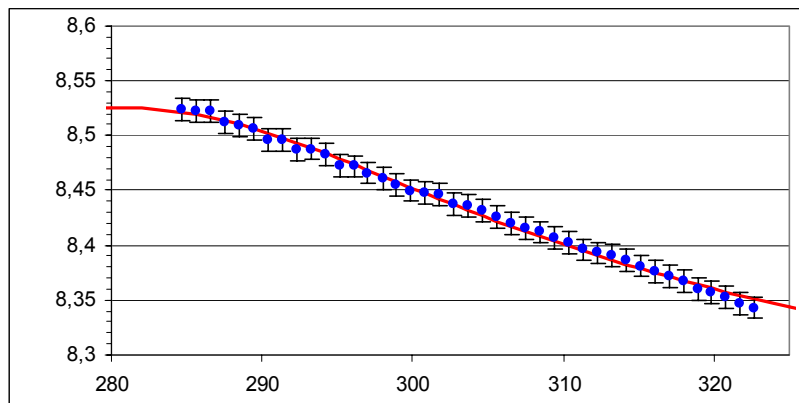
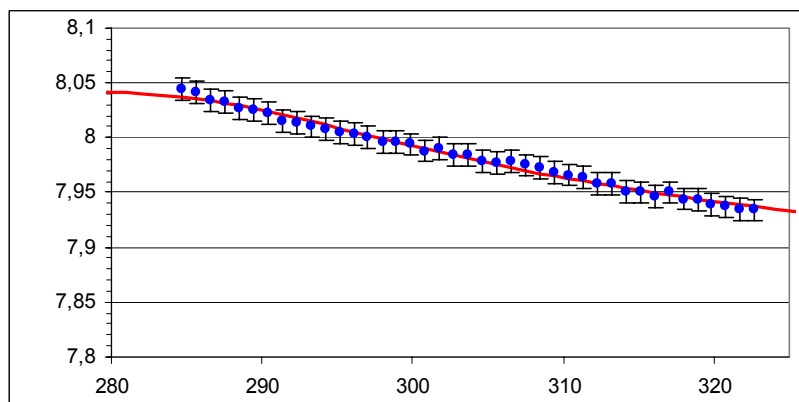
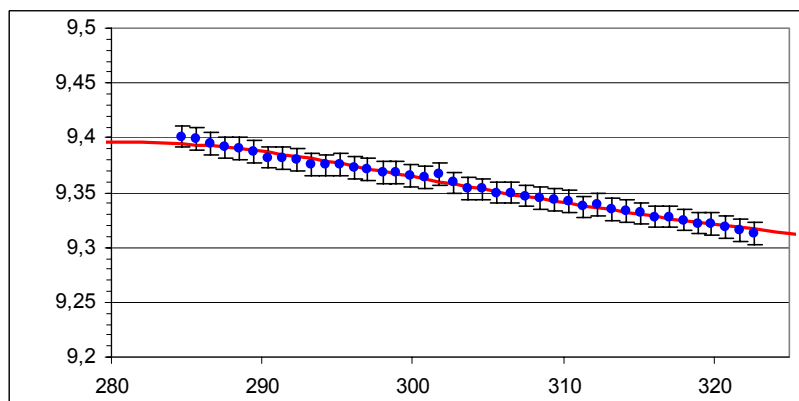
$H^N 40$  $H^N 42$  $H^N 43$ 

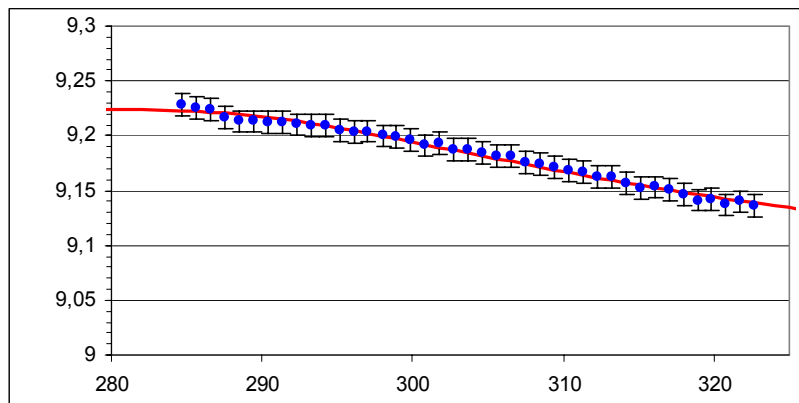
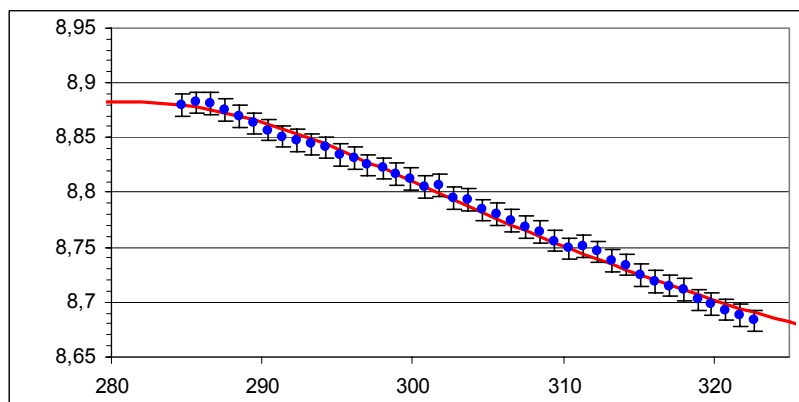
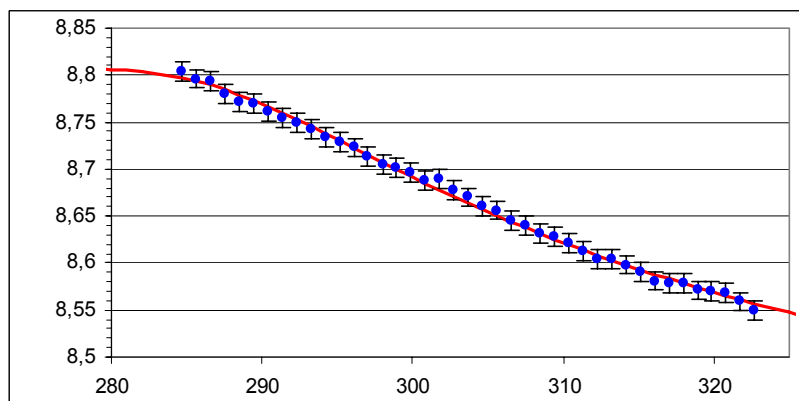
**H<sup>N</sup> 45****H<sup>N</sup> 46****H<sup>N</sup> 49**

$H^N$  50 $H^N$  51 $H^N$  52

**H<sup>N</sup> 53****H<sup>N</sup> 59****H<sup>N</sup> 60**

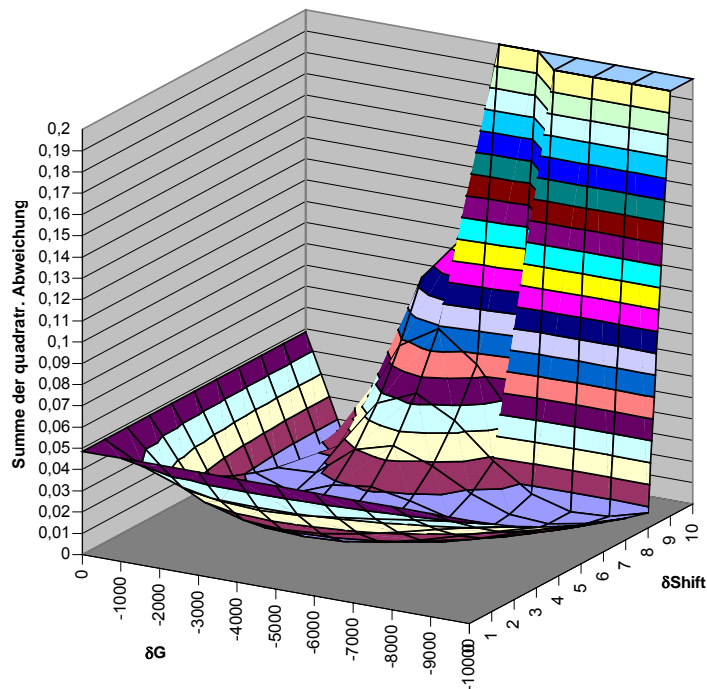


$H^N 64$  $H^N 71$  $H^N 75$ 

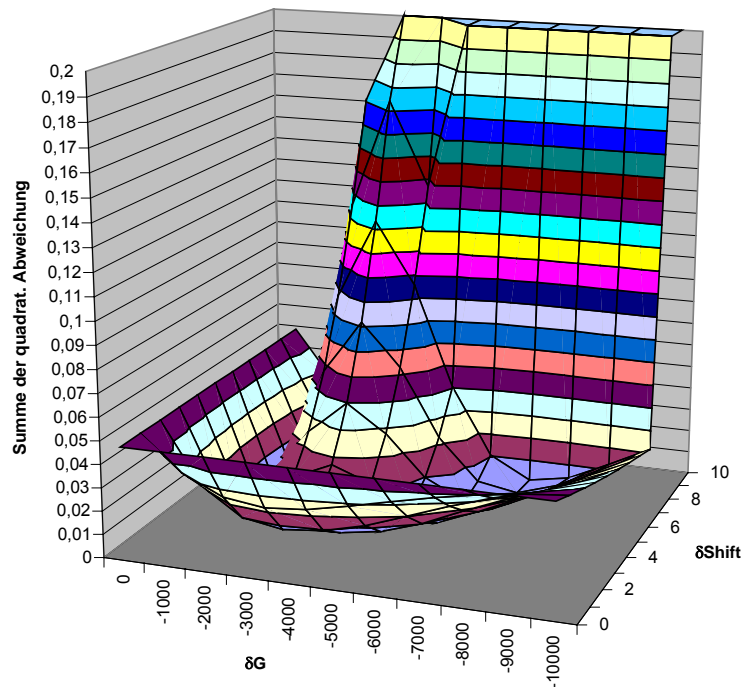
**H<sup>N</sup> 78****H<sup>N</sup> 79****H<sup>N</sup> 83**

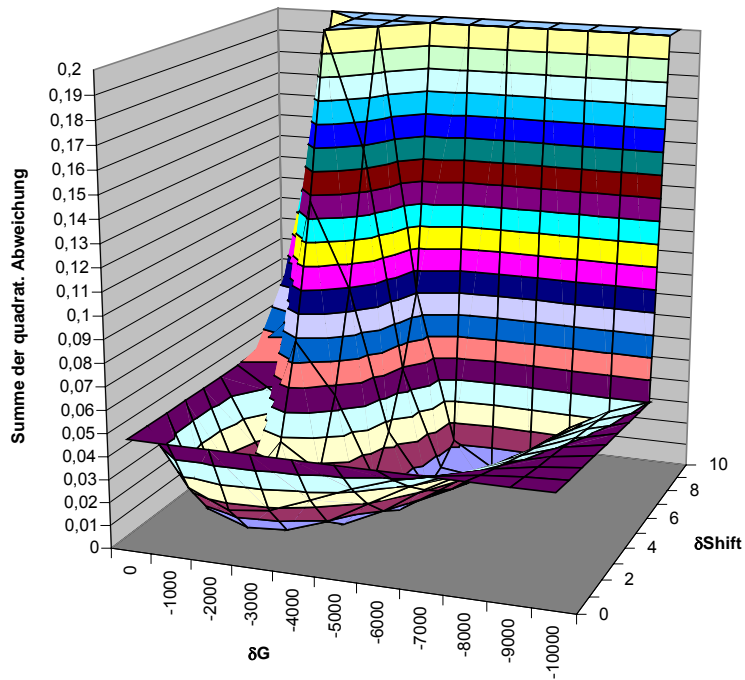
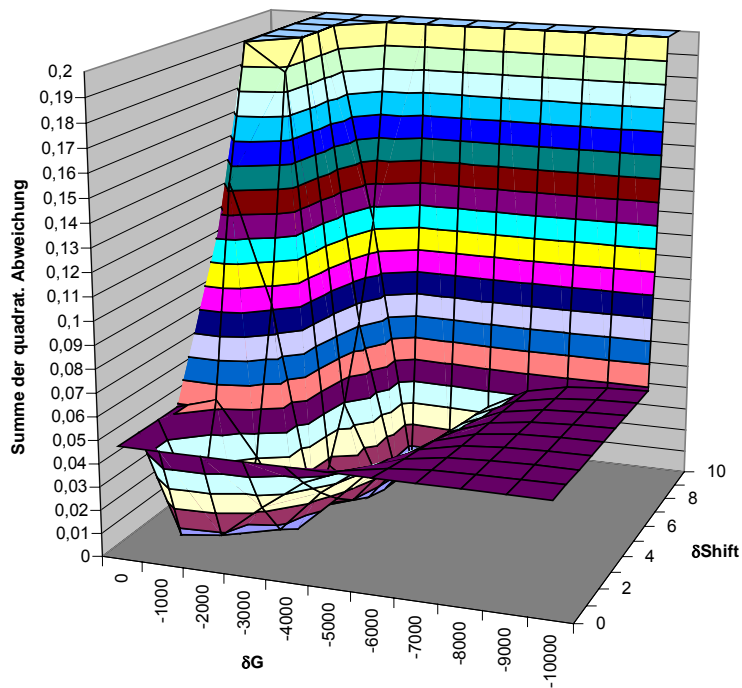
## 9.3. Untersuchungen zu unterschiedlichen $T_0$ Werten

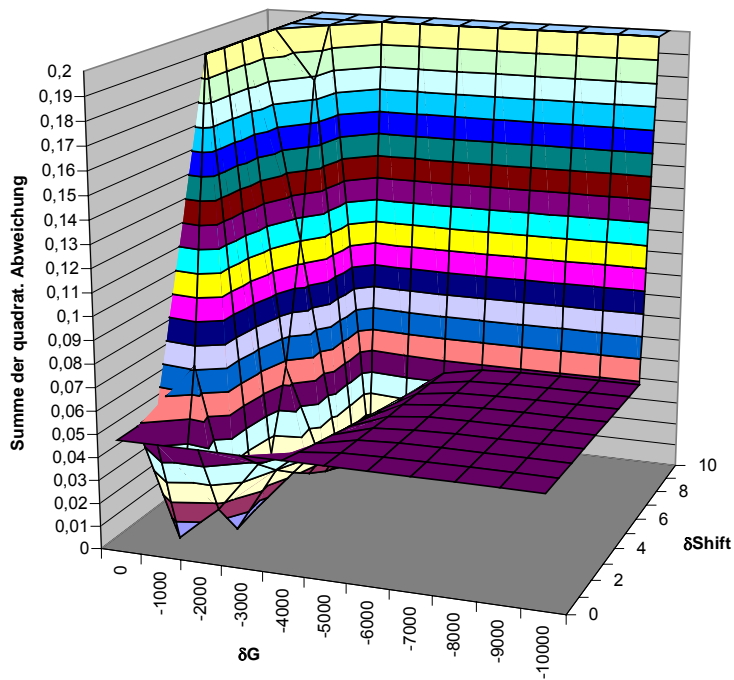
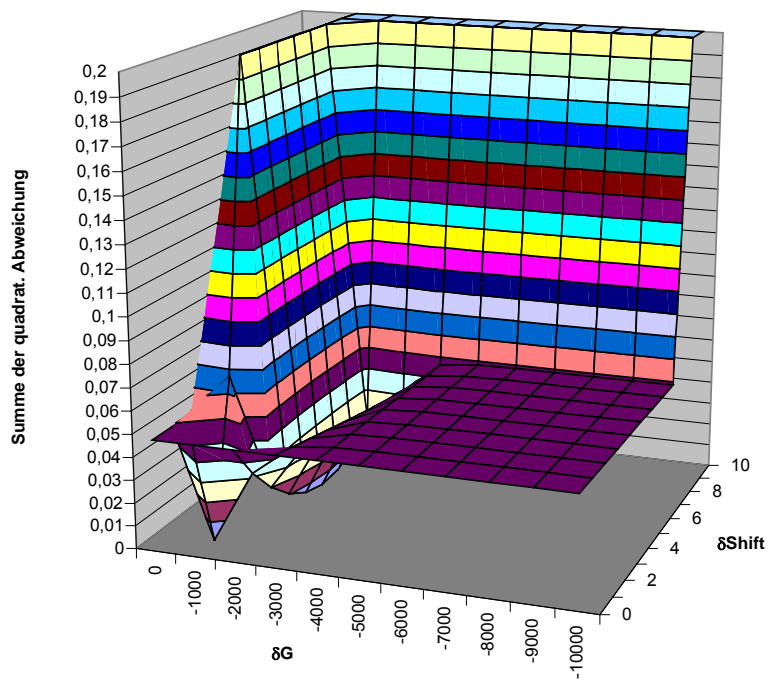
Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=0K$

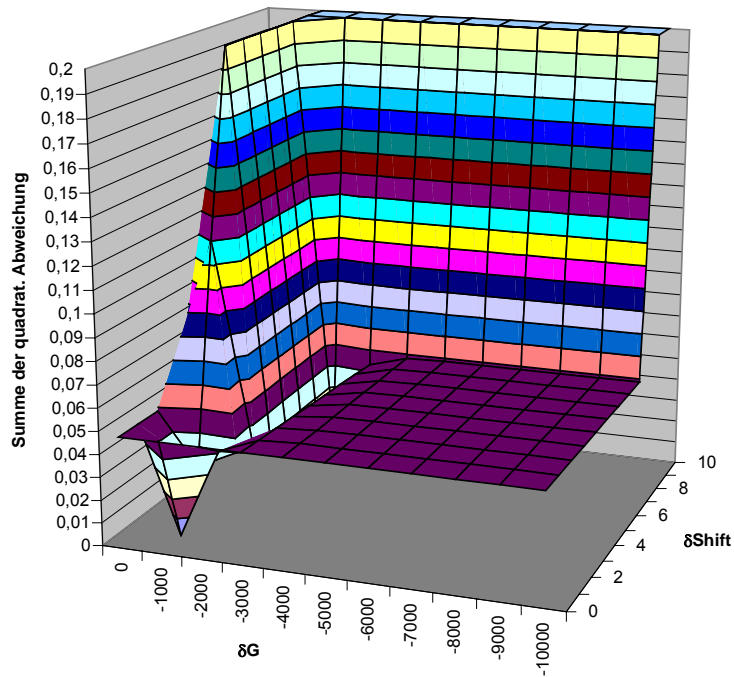
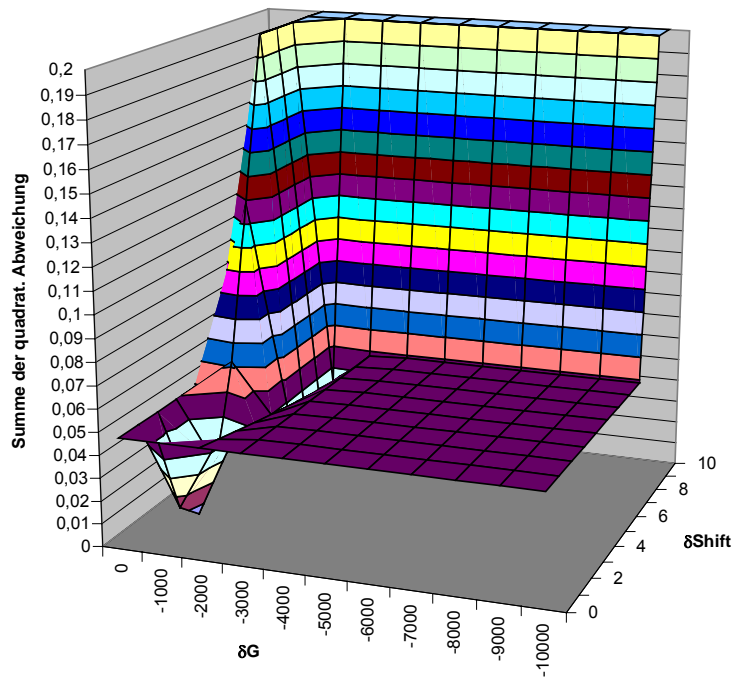


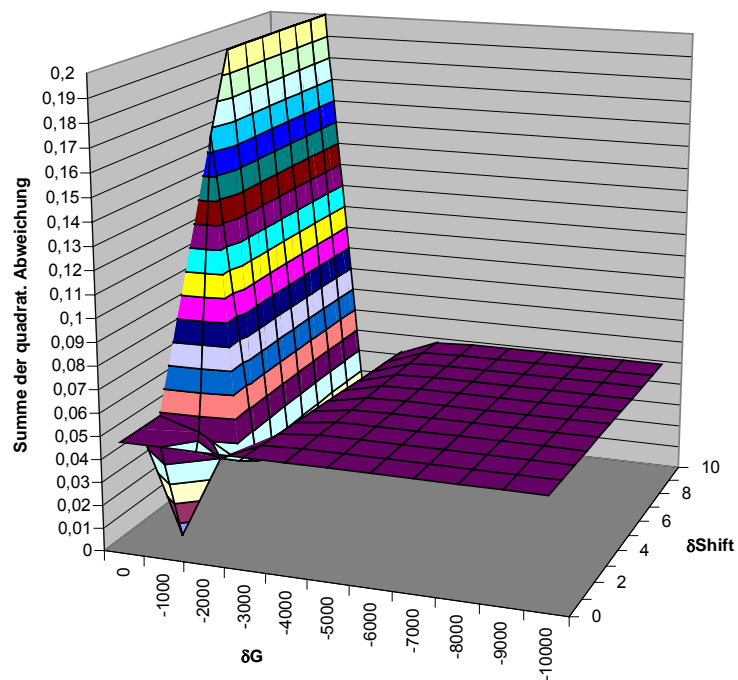
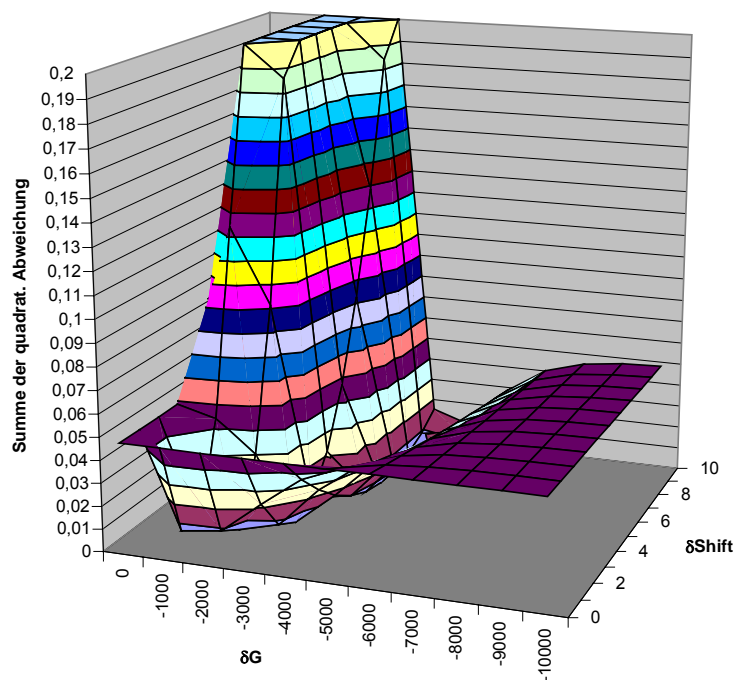
Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=70K$



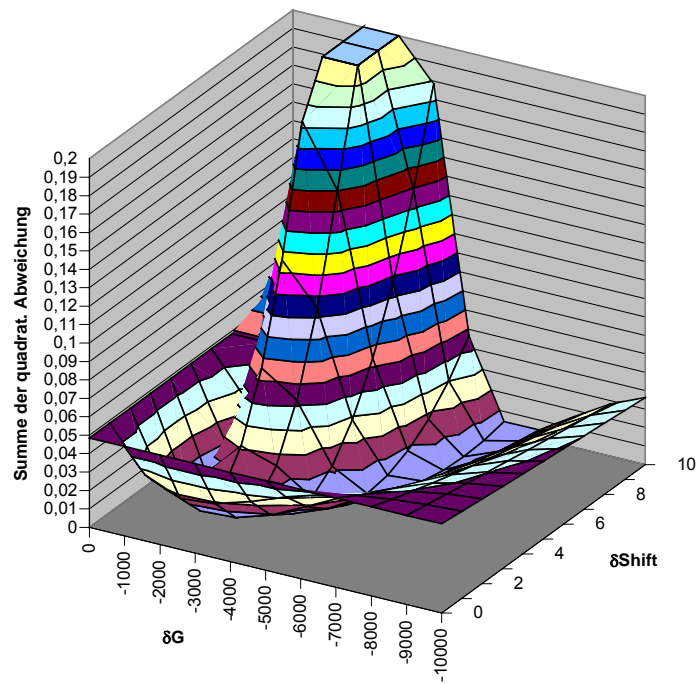
Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=135\text{K}$ Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=200\text{K}$ 

Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=240\text{K}$ Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=260\text{K}$ 

Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=270\text{K}$ Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=280\text{K}$ 

Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=340\text{K}$ Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=410\text{K}$ 

Quadratische Abweichung bei unterschiedlichen Kombinationen von  $\delta\text{Shift}$  und  $\delta G$  bei  $T_0=480\text{K}$





---

# Danksagung

Die vorliegende Arbeit wurde im Zeitraum von April 2001 bis April 2005 im Institut für Biophysik und Physikalische Biochemie der Universität Regensburg unter Anleitung von Prof. Dr. Dr. H. R. Kalbitzer angefertigt. Viele Kollegen waren am Entstehen dieser Arbeit direkt und indirekt beteiligt. Ihnen sei an dieser Stelle gedankt.

Zuerst möchte ich mich bei Prof. Dr. Dr. H. R. Kalbitzer bedanken, der mit seiner freundlichen und umgänglichen Art stets Interesse an meiner Arbeit zeigte, und mit interessanten Anregungen und Vorschlägen viel zum Gelingen dieser Arbeit beigetragen hat. Mein besonderer Dank gilt Dr. Wolfram Gronwald, der mich mit großem Sachverstand betreute, alle Fragen und Probleme gewissenhaft untersuchte und der trotz mangelnder Zeit immer ein offenes Ohr für alle Anliegen hatte.

Weiterhin möchte ich mich bei allen, auch namentlich nicht genannten, Lehrstuhlangehörigen für ihre Kollegialität und Freundschaft bedanken, die das Arbeiten hier wahrhaft angenehm machte.

Dabei seien besonders meine "Mitinsassen" Barbara Domogalla und Ralph Elsner genannt, die immer gut für ein Gespräch, eine Tasse Tee, oder einen kleinen Scherz waren.

Für die gute Zusammenarbeit bei programmiertechnischen Fragen, und bei Problembewältigungsstrategien im Zusammenhang mit AUREMOL und anderen Widrigkeiten des Lebens möchte ich Konrad Brunner, Jochen Trenner und Adel Nasser danken.

Außerdem danke ich Dr. Werner Kremer dafür, dass ich seine Protein-Strukturen für meine Programmtests verwenden durfte, ohne dass er zunächst wusste, ob das gut oder schlecht für ihn sein würde und Prof. Dr. Eike Brunner für seine immerwährende Freundlichkeit und die vorbildlich rechtzeitige Bereitstellung von Vorlesungsankündigungen und Seminaren für die Lehrstuhlhomepage.

Herrn Prof. Dr. Lang danke ich für seine ruhige, besonnene und freundliche Art, die schon während des Studiums maßgeblich dafür sorgte, mein Interesse an der Biophysik zu wecken.

Zu guter Letzt möchte ich Fritz herzlich dafür danken, dass er mich während der Doktorarbeit und der Prüfungsvorbereitung ertragen und immer wieder aufgebaut hat, und auch fachliche Diskussionen zu unmöglichen Zeiten und wechselnden Orten immer geduldig über sich ergehen ließ.

---

---

# Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig angefertigt, und keine Hilfsmittel, außer den angegebenen, benutzt habe.

Saal a. d. Donau, April 2005

---

(Renate Kirchhöfer)

---